

Transfection efficiency can be monitored using marker genes, such as green fluorescent protein, encoded by the same vector as the TAP genes. Cells expressing equal levels of green fluorescent protein but the highest levels of MHC class I molecules, as a marker of efficient TAP genes, are then sorted using flow cytometry, and the evolved TAP genes are then recovered from these cells by, for example, PCR or by recovering the entire vectors.

These sequences can then subjected into new rounds of reassembly (optionally in combination with other directed evolution methods described herein), selection and recovery, if further optimization is desired. Molecular evolution of TAP genes can be combined with simultaneous evolution of the desired antigen. Simultaneous evolution of the desired antigen can further improve the efficacy of presentation of antigenic peptides following DNA vaccination. The antigen can be evolved, using polynucleotide reassembly (optionally in combination with other directed evolution methods described herein), to contain structures that allow optimal presentation of desired antigenic peptides when optimal TAP genes are expressed. TAP genes that are optimal for presentation of antigenic peptides of one given antigen may be different from TAP genes that are optimal for presentation of antigenic peptide of another antigen. Polynucleotide (e.g. gene, promoter, enhancer, intron, & the like) reassembly (optionally in combination with other directed evolution methods described herein) technique is ideal, and perhaps the only, method to solve this type of problems. Efficient presentation of desired antigenic peptides can be analyzed using specific cytotoxic T lymphocytes, for example, by measuring the cytokine production or CTL activity of the T lymphocytes using methods known to those of skill in the art.

2.7.3. CYTOTOXIC T-CELL INDUCING SEQUENCES AND IMMUNOGENIC AGONIST SEQUENCES

Certain proteins are better able than others to carry MHC class I epitopes
5 because they are more readily used by the cellular machinery involved in the necessary processing for class I epitope presentation. The invention provides methods of identifying expressed polypeptides that are particularly efficient in traversing the various biosynthetic and degradative steps leading to class I epitope presentation and the use of these polypeptides to enhance presentation of CTL
10 epitopes from other proteins.

In one embodiment, the invention provides Cytotoxic T-cell Inducing Sequences (CTIS), which can be used to carry heterologous class I epitopes for the purpose of vaccinating against the pathogen from which the heterologous
15 epitopes are derived. One example of a CTIS is obtained from the hepatitis B surface antigen (HBsAg), which has been shown to be an effective carrier for its own CTL epitopes when delivered as a protein under certain conditions. DNA immunization with plasmids expressing the HBsAg also induces high levels of CTL activity. The invention provides a shorter, truncated fragment of the HBsAg
20 polypeptide which functions very efficiently in inducing CTL activity, and attains CTL induction levels that are higher than with the HBsAg protein or with the plasmids encoding the full-length HBsAg polypeptide. Synthesis of a CTIS derived from HBsAg is described in Example 3; and a diagram of a CTIS is shown, described &/or referenced herein (including incorporated by reference).
25

The ER localization of the truncated polypeptide may be important in achieving suitable proteolytic liberation of the peptide(s) containing the CTL epitopes (see Cresswell � Craiu et al. (1997) Proc. Nat'l. Acad. Sei. USA 94: 10850-10855). The preS2 region and the transmembrane region provide T-

helper epitopes which may be important for the induction of a strong cytotoxic immune response. Because the truncated CTIS polypeptide has a simple structure, it is possible to attach one or more heterologous class I epitope sequences to the C-terminal end of the polypeptide without having to maintain any specific protein conformation. Such sequences are then available to the class I epitope processing mechanisms. The size of the polypeptide is not subject to the normal constraints of the native HBsAg structure. Therefore the length of the heterologous sequence and thus the number of included CTL epitopes is flexible. This is shown schematically herein. The ability to include a long sequence containing either multiple and distinct class I sequences, or alternatively different variations of a single CTL sequence, allows stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methodology to be applied.

The invention also provides methods of obtaining Immunogenic Agonist Sequences (IAS) which induce CTLs capable of specific lysis of cells expressing the natural epitope sequence. In some cases, the reactivity is greater than if the CTL response is induced by the natural epitope. Such IAS-induced CTL may be drawn from a T-cell repertoire different from that induced by the natural sequence. In this way, poor responsiveness to a given epitope can be overcome by recruiting T cells from a larger pool. In order to discover such IAS, the amino acid at each position of a CTL-inducing peptide (excluding perhaps the positions of the so-called anchor residues) can be varied over the range of the 19 amino acids not normally present at the position. stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methodology can be used to scan a large range of sequence possibilities.

A synthetic gene segment containing multiple copies of the original epitope sequence can be prepared such that each copy possesses a small number

of nucleotide changes. The gene segment can be experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) to create a diverse range of CTL epitope sequences, some of which should function as IAS. This process is illustrated herein.

5

In practice, oligonucleotides are typically constructed in accordance with the above design and polymerized enzymatically to form the synthetic gene segment of the concatenated epitopes. Restriction sites can be incorporated into a fraction of the oligonucleotides to allow for cleavage and selection of given size ranges of the concatenated epitopes, most of which will have different sequences and thus will be potential IAS. The epitope-containing gene segment can be joined by appropriate cloning methods to a CTIS, such as that of HBsAg. The resulting plasmid constructions can be used for DNA-based C immunization and CTL induction.

15

2.8. GENETIC VACCINE PHARMACEUTICAL COMPOSITIONS AND METHODS OF ADMINISTRATION

20

Using genetic vaccines in prophylaxis and therapy of infectious diseases, autoimmune diseases, other inflammatory conditions, allergies, asthma, and cancer and the prevention of metastasis

25

The vector components and multicomponent genetic vaccines of the invention are useful for treating and/or preventing various diseases and other conditions. For example, genetic vaccines that employ the reagents obtained according to the methods of the invention are useful in both prophylaxis and therapy of infectious diseases, including those caused by any bacterial, fungal,

viral, or other pathogens of mammals. The reagents obtained using the invention can also be used for treatment of autoimmune diseases including, for example, rheumatoid arthritis, SLE, diabetes mellitus, myasthenia gravis, reactive arthritis, ankylosing spondylitis, and multiple sclerosis. These and other inflammatory
5 conditions, including IBD, psoriasis, pancreatitis, and various immunodeficiencies, can be treated using genetic vaccines that include vectors and other components obtained using the methods of the invention. Genetic vaccine vectors and other reagents obtained using the methods of the invention can be used to treat allergies and asthma. Moreover, the use of genetic vaccines
10 have great promise for the treatment of cancer and prevention of metastasis. By inducing an immune response against cancerous cells, the body's immune system can be enlisted to reduce or eliminate cancer.

15 Use of Recombinant Multivalent Antigens

The multivalent antigens of the invention are useful for treating and/or preventing the various diseases and conditions with which the respective antigens are associated. For example, the multivalent antigens can be expressed in a
20 suitable host cell and are administered in polypeptide form. Suitable formulations and dosage regimes for vaccine delivery are well known to those of skill in the art. The improved immunomodulatory polynucleotides and polypeptides of the invention are useful for treating and/or preventing the various diseases and conditions with which the respective antigens are associated.

25

An antigen for a particular condition can be optimized using reassembly (&/or one or more additional directed evolution methods described herein) and selection methods analogous to those described herein.

5

In presently preferred embodiments, the reagents obtained using the invention (e.g. optimized experimentally generated polynucleotides that encode improved allergens), are used in conjunction with a genetic vaccine. The choice of vector and components can also be optimized for the particular purpose of treating allergy or other conditions. In presently preferred embodiments, the optimized genetic vaccine components are used in conjunction with other optimized genetic vaccine reagents. For example, an antigen that is useful for a particular condition can be optimized by methods analogous to the reassembly (&/or one or more additional directed evolution methods described herein) and screening methods described herein.

15

The polynucleotide that encodes the recombinant antigenic polypeptide can be placed under the control of a promoter, e.g., a high activity or tissue-specific promoter. The promoter used to express the antigenic polypeptide can itself be optimized using reassembly (&/or one or more additional directed evolution methods described herein) and selection methods analogous to those described herein., as described in International Application No. PCTIUS97/17300 (International Publication No. WO 98/13487).

20

The vector can contain immunostimulatory sequences such as are described herein. A vector engineered to direct a T_H1 response is preferred for many of the immune responses mediated by the antigens described herein. The reagents obtained using the methods of the invention can also be used in conjunction with multicomponent genetic vaccines, which are capable of tailoring

25

an immune response as is most appropriate to achieve a desired effect. It is sometimes advantageous to employ a genetic vaccine that is targeted for a particular target cell type (e.g., an antigen presenting cell or an antigen processing cell); suitable targeting methods are described herein.

5

Delivery of genetic vaccines and delivery vehicles to mammals *in vivo* and *ex vivo*

10 Genetic vaccines, (e.g. genetic vaccines that include the optimized experimentally generated polynucleotides obtained as described herein, such as genetic vaccines that encode the multivalent antigens described herein, including the multicomponent genetic vaccines described herein), can be delivered to a mammal (including humans) to induce a therapeutic or prophylactic immune
15 response. Vaccine delivery vehicles can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, intracranial, anal, vaginal, oral, buccal route or they can be inhaled) or they can be administered by topical application.

20 Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

25

Delivery methods and references

- A large number of delivery methods are well known to those of skill in the art. Such methods include, for example liposome-based gene delivery (Debs and Zhu (1993) WO 93/24640; Mannino and Gould-Fogerite (1988) BioTechniques 6(7): 682- 691; Rose U.S. Pat No. 5,279,833; Brigham (1991) WO 91/06309; and Felgner et al. (1987) Proc. Natl. Acad. Sci. USA 84: 7413-7414), as well as use of viral vectors (e.g., adenoviral (see, e.g., Berns et al. (1995) Ann. NY Acad Sci. 772: 95-104; Ali et al. (1994) Gene Ther. 1: 367-3 84; and Haddada et al. (1995) Curr. Top. Microbiol. Immunol. 199 (Pt 3): 297- 306 for review), papillomaviral, retroviral (see, e.g., Buchscher et al. (1992) J Virol. 66(5) 2731-2739; Johann et al. (1992) J Virol. 66 (5):163 5-1640 (1992); Sommerfelt et al. , (1990) Virol. 176:58-59; Wilson et al. (1989) J Virol. 63:2374-2378; Miller et al., J Virol. 65:2220-2224 (1991); Wong-Staal et al., PCT/US94/05700, and Rosenberg and Fauci (1993) in Fundamental Immunology, Third Edition Paul (ed) Raven Press, Ltd., New York and the references therein, and Yu et al., Gene Therapy (1994) supra.), and adeno-associated viral vectors (see, West et al. (1987) Virology 160:3 8-47; Carter et al. (1989) U. S. Patent No. 4,797,3 68; Carter et al. WO 93/24641 (1993); Kotin (1994) Human Gene Therapy 5:793 - 801; Muzyczka (1994) J Clin. Invest. 94:1351 and Samulski (supra) for an overview of AAV vectors; see also, Lebkowski, U.S. Pat. No. 5,173,414; Tratschin et al. (1985) Mol. Cell. Biol. 5(11):3251-3260; Tratschin, et al. (1984) Mol. Cell. Biol., 4:2072- 2081; Hermonat and Muzyczka (1984) Proc. Natl. Acad Sci. USA, 81:6466-6470; McLaughlin et al. (1988) and Samulski et al. (1989) J Virol., 63:03 822-3 828), and the like.

Introduction of "Naked" DNA and/or RNA that comprises a genetic vaccine directly into a tissue or using "biolistic" or particle-mediated transformation, both *in vivo* and *ex vivo*

5 "Naked" DNA and/or RNA that comprises a genetic vaccine can be introduced directly into a tissue, such as muscle. See, e.g., USPN 5,580, 859. Other methods such as "biolistic" or particle-mediated transformation (see, e.g., Sanford et al., USPN 4,945,050; USPN 5,036,006) are also suitable for introduction of genetic vaccines into cells of a mammal according to the
10 invention. These methods are useful not only for *in vivo* introduction of DNA into a mammal, but also for *ex vivo* modification of cells for reintroduction into a mammal. As for other methods of delivering genetic vaccines, if necessary, vaccine administration is repeated in order to maintain the desired level of immunomodulation.

15

Methods of administering packaged nucleic acids in mammals for transduction of cells *in vivo*

Genetic vaccine vectors (e.g., adenoviruses, liposomes, papillornaviruses, retroviruses, etc.) can be administered directly to the mammal for transduction of cells *in vivo*. The genetic vaccines obtained using the methods of the invention can be formulated as pharmaceutical compositions for administration in any suitable manner, including parenteral (e.g., subcutaneous, intramuscular, intradermal, or intravenous), topical, oral, rectal, intrathecal, buccal (e.g., sublingual), or local administration, such as by aerosol or transdermally, for prophylactic and/or therapeutic treatment. Pretreatment of skin, for example, by use of hair-removing agents, may be useful in transdermal delivery. Suitable methods of administering such packaged nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention. A variety of aqueous carriers can be used, e.g., buffered saline and the like. These solutions are sterile and generally free of undesirable matter. These compositions may be sterilized by conventional, well known sterilization techniques. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH adjusting and buffering agents, toxicity adjusting agents and the like, for example, sodium acetate, sodium chloride, potassium chloride, calcium chloride, sodium lactate and the like. The concentration of genetic vaccine vector in these formulations

can vary widely, and will be selected primarily based on fluid volumes, viscosities, body weight and the like in accordance with the particular mode of administration selected and the patient's needs.

- 5 Formulations suitable for oral administration can consist of (a) liquid solutions, such as an effective amount of the packaged nucleic acid suspended in diluents, such as water, saline or PEG 400; (b) capsules, sachets or tablets, each containing a predetermined amount of the active ingredient, as liquids, solids, granules or gelatin; (c) suspensions in an appropriate liquid; and (d) suitable
10 emulsions. Tablet forms can include one or more of lactose, sucrose, mannitol, sorbitol, calcium phosphates, corn starch, potato starch, tragacanth, microcrystalline cellulose, acacia, gelatin, colloidal silicon dioxide, croscannellose sodium, talc, magnesium stearate, stearic acid, and other excipients, colorants, fillers, binders, diluents, buffering agents, moistening
15 agents, preservatives, flavoring agents, dyes, disintegrating agents, and pharmaceutically compatible carriers.

- Lozenge forms can comprise the active ingredient in a flavor, usually sucrose and acacia or tragacanth, as well as pastilles comprising the active
20 ingredient in an inert base, such as gelatin and glycerin or sucrose and acacia emulsions, gels, and the like containing, in addition to the active ingredient, carriers known in the art. It is recognized that the genetic vaccines, when administered orally, must be protected from digestion. This is typically accomplished either by complexing the vaccine vector with a composition to
25 render it resistant to acidic and enzymatic hydrolysis or by packaging the vector in an appropriately resistant carrier such as a liposome. Means of protecting vectors from digestion are well known in the art. The pharmaceutical compositions can be encapsulated, e. g., in liposomes, or in a formulation that provides for slow release of the active ingredient.

The packaged nucleic acids, alone or in combination with other suitable components, can be made into aerosol formulations (e.g., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into

5 pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like. Suitable formulations for rectal administration include, for example, suppositories, which consist of the packaged nucleic acid with a suppository base. Suitable suppository bases include natural or synthetic triglycerides or paraffin hydrocarbons. In addition, it is also possible to use gelatin

10 rectal capsules which consist of a combination of the packaged nucleic acid with a base, including, for example, liquid triglycerides, polyethylene glycols, and paraffin hydrocarbons.

Formulations suitable for parenteral administration, such as, for example,

15 by intraarticular (in the joints), intravenous, intramuscular, intradermal, intraperitoneal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can

20 include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally.

Parenteral administration and intravenous administration are the preferred methods of administration

5 The formulations of packaged nucleic acid can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. Cells transduced by the packaged nucleic acid can also be administered intravenously or parenterally.

10

Dose size

15 The dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial therapeutic response in the patient over time. The dose will be determined by the efficacy of the particular vector employed and the condition of the patient, as well as the body weight or vascular surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that
20 accompany the administration of a particular vector, or transduced cell type in a particular patient.

25 In determining the effective amount of the vector to be administered in the treatment or prophylaxis of an infection or other condition, the physician evaluates vector toxicities, progression of the disease, and the production of anti-vector antibodies, if any. In general, the dose equivalent of a naked nucleic acid from a vector is from about 1 μ g to 1 mg for a typical 70 kilogram patient, and doses of vectors used to deliver the nucleic acid are calculated to yield an

equivalent amount of therapeutic nucleic acid. Administration can be accomplished via single or divided doses.

5 In therapeutic applications, compositions are administered to a patient suffering from a disease (e.g., an infectious disease or autoimmune disorder) in an amount sufficient to cure or at least partially arrest the disease and its complications. An amount adequate to accomplish this is defined as a "therapeutically effective dose." Amounts effective for this use will depend upon the severity of the disease and the general state of the patient's health. Single or
10 multiple administrations of the compositions may be administered depending on the dosage and frequency as required and tolerated by the patient. In any event, the composition should provide a sufficient quantity of the proteins of this invention to effectively treat the patient.

15 In prophylactic applications, compositions are administered to a human or other mammal to induce an immune response that can help protect against the establishment of an infectious disease or other condition.

20 **Ability to determine toxicity therapeutic efficacy**

The toxicity and therapeutic efficacy of the genetic vaccine vectors provided by the invention are determined using standard pharmaceutical procedures in cell cultures or experimental animals. One can determine the LD₅₀
25 (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population) using procedures presented herein and those otherwise known to those of skill in the art.

More on dosage

A typical pharmaceutical composition for intravenous administration would be about 0.1 to 10 mg per patient per day. Dosages from 0.1 up to about
5 100 mg per patient per day may be used, particularly when the drug is administered to a secluded site and not into the blood stream, such as into a body cavity or into a lumen of an organ. Substantially higher dosages are possible in topical administration. Actual methods for preparing parenterally administrable compositions will be known or apparent to those skilled in the art and are
10 described in more detail in such publications as Remington's Pharmaceutical Science, 15th ed., Mack Publishing Company, Easton, Pennsylvania (1980).

Packaging/dispenser devices

15

The genetic vaccines obtained using the methods of the invention (e.g. the multivalent antigenic polypeptides of the invention, and genetic vaccines that express the polypeptides) can be packaged in packs, dispenser devices, and kits for administering genetic vaccines to a mammal. For example, packs or dispenser
20 devices that contain one or more unit dosage forms are provided. Typically, instructions for administration of the compounds will be provided with the packaging, along with a suitable indication on the label that the compound is suitable for treatment of an indicated condition. For example, the label may state that the active compound within the packaging is useful for treating a particular
25 infectious disease, autoimmune disorder, tumor, or for preventing or treating other diseases or conditions that are mediated by, or potentially susceptible to, a mammalian immune response.

2.9. USES OF GENETIC VACCINES

Genetic vaccines which include optimized vector modules and other reagents provided by the invention are useful for treating many diseases and other conditions that are either mediated by a mammalian immune system or are susceptible to treatment by an appropriate immune response. Representative examples of these diseases & antigens appropriate for each are listed below, described herein, or incorporated by reference.

10

Substrates for evolution of optimized recombinant antigens

The invention provides methods of obtaining experimentally generated polynucleotides that encode antigens that exhibit improved ability to induce an immune response to a pathogenic agent. The methods are applicable to a wide range of pathogenic agents, including potential biological warfare agents and other organisms and polypeptides that can cause disease and toxicity in humans and other animals. The following examples are merely illustrative, and not limiting.

20

2.9.1. INFECTIOUS DISEASES

Genetic vaccine vectors obtained according to the methods of the invention are useful in both prophylaxis and therapy of infectious diseases, including those caused by any bacterial, fungal, viral, or other pathogens of mammals. In some embodiments, protection is conferred by use of a genetic vaccine vector that will express an antigen (either or both of a humoral antigen or a T cell antigen) of the pathogen of interest. In preferred embodiments, the

25

antigen is evolved using the methods of the invention in order to obtain optimized antigens as described herein. The vector induces an immune response against the antigen. One or several antigens or antigen fragments can be included in one genetic vaccine delivery vehicle. Examples of pathogens and corresponding polypeptides from which an antigen can be obtained include, but are not limited to, HIV (gp120, gp160), hepatitis B, C, D, E (surface antigen), rabies (glycoprotein), *Schistosoma mansoni* (calpain; Jankovic (1996) J Immunol. 157: 806-14). Other pathogen infections that are treatable using genetic vaccine vectors include, for example, herpes zoster, herpes simplex -1 and -2, tuberculosis (including chronic, drug-resistant), lyme disease (*Borrelia burgdorferii*), syphilis, parvovirus, rabies, human papillomavirus, and the like.

2.9.1.1 BACTERIAL PATHOGENS AND TOXINS

In some embodiments, the methods of the invention are applied to bacterial pathogens, as well as to toxins produced by bacteria and other organisms. One can use the methods to obtain experimentally generated polypeptides that can induce an immune response against the pathogen, as well as recombinant toxins that are less toxic than native toxin polypeptides. Often, the polynucleotides of interest encode polypeptides that are present on the surface of the pathogenic organism. Among the pathogens for which the methods of the invention are useful for producing protective immunogenic experimentally generated polypeptides are the *Yersinia* species.

Yersinia pestis, the causative agent of plague, is one of the most virulent bacteria known with LD₅₀ values in mouse of less than 10 bacteria. The pneumonic form of the disease is readily spread between humans by aerosol or infectious droplets and can be lethal within days. A particularly preferred target

for obtaining a experimentally generated polypeptide that can protect against Yersinia infection is the V antigen, which is a 37 kDa virulence factor, induces protective immune responses and is currently being evaluated as a subunit vaccine (Brubaker (1991) Current Investigations of the Microbiology of Yersinae, 12: 127). The V-antigen alone is not toxic, but Y pestis isolates that lack the V-antigen are avirulent. The Yersinia V- antigen has been successfully produced in E. coli by several groups (Leary et al. (1995) Infect. Immun. 3: 2854). Antibodies that recognize the V-antigen can provide passive protection against homologous strains, but not against heterologous strains. Similarly, immunization with purified V antigen protects against only homologous strains. To obtain cross-protective recombinant V antigen, in a preferred embodiment, V antigen genes from various Yersinia species are subjected to polynucleotide reassembly (optionally in combination with other directed evolution methods described herein). The genes encoding the V antigen from Y. pestis, Y. enterocolitica, and Y. pseudotuberculosis, for example, are 92-99% identical at the DNA level, making them ideal for optimization using family reassembly (optionally in combination with other directed evolution methods described herein) according to the methods of the invention. After reassembly (optionally in combination with other directed evolution methods described herein), the library of recombinant nucleic acids is screened and/or selected for those that encode recombinant V antigen polypeptides that can induce an improved immune response and/or have greater cross- protectivity.

Bacillus anthracis, the causative agent of anthrax, is another example of a bacterial target against which the methods of the invention are useful. The anthrax protective antigen (PA) provides protective immune responses in test animals, and antibodies against PA also provide some protection. However, the immunogenicity of PA is relatively poor, so multiple injections are typically required when wild-type PA is used. Co- vaccination with lethal factor (LF) can

improve the efficacy of wild-type PA vaccines, but toxicity is a limiting factor. Accordingly the stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly and antigen library immunization methods of the invention can be used to obtain nontoxic LF. Polynucleotides that
5 encode LF from various B. anthracis strains are subjected to family reassembly (optionally in combination with other directed evolution methods described herein). The resulting library of recombinant LF nucleic acids can then be screened to identify those that encode recombinant LF polypeptides that exhibit reduced toxicity. For example, one can inoculate tissue culture cells with the
10 recombinant LF polypeptides in the presence of PA and select those clones for which the cells survive. If desired, one can then backcross the nontoxic LF polypeptides to retain the immunogenic epitopes of LF. Those that are selected through the first screen can then be subjected to a secondary screen. For example, one can test for the ability of the recombinant nontoxic LF polypeptides to induce
15 an immune response (e.g., CTL or antibody response) in a test animal such as mice. In preferred embodiments, the recombinant nontoxic LF polypeptides are then tested for ability to induce protective immunity in test animals against challenge by different strains of B. anthracis.

20 The protective antigen (PA) of B. anthracis is also a suitable target for the methods of the invention. PA-encoding nucleic acids from various strains of B. anthracis are subjected to stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly. One can then screen for proper folding in, for example, E. coli, using polyclonal antibodies. Screening for
25 ability to induce broad- spectrum antibodies in a test animal is also typically used, either alone or in addition to a preliminary screening method. In presently preferred embodiments, those experimentally generated polynucleotides that exhibit the desired properties can be backcrossed so that the immunogenic epitopes are maintained. Finally, the selected recombinants are tested for ability

to induce protective immunity against different strains of *B. anthracis* in a test animal.

- The *Staphylococcus aureus* and *Streptococcus* toxins are another example
- 5 of a target polypeptide that can be altered using the methods of the invention. Strains of *Staphylococcus aureus* and group A *Streptococci* are involved in a range of diseases, including food poisoning, toxic shock syndrome, scarlet fever and various autoimmune disorders. They secrete a variety of toxins, which include at least five cytolytic toxins, a coagulase, and a variety of enterotoxins.
- 10 The enterotoxins are classified as superantigens in that they crosslink MHC class II molecules with T cell receptors to cause a constitutive T cell activation (Fields et al. (1996) *Nature* 384: 188). This results in the accumulation of pathogenic levels of cytokines that can lead to multiple organ failure and death. At least thirty related, yet distinct enterotoxins have been sequenced and can be phylogenetically
- 15 grouped into families. Crystal structures have been obtained for several members alone and in complex with MHC class II molecules. Certain mutations in the MHC class II binding site of the toxins strongly reduce their toxicity and can form the basis of attenuated vaccines (Woody et al. (1997) *Vaccine* 15: 133). However, a successful immune response to one type of toxin may provide protection against
- 20 closely related family members, whereas little protection against toxins from the other families is observed. Family reassembly (optionally in combination with other directed evolution methods described herein) of enterotoxin genes from various family members can be used to obtain recombinant toxin molecules that have reduced toxicity and can induce a cross-protective immune response.
- 25 Experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens can also be screened to identify antigens that elicit neutralizing antibodies in an appropriate animal model such as mouse or monkey. Examples of such assays can include ELISA formats in which the elicited antibodies prevent binding of the enterotoxin to the MHC complex and/or

T cell receptors on cells or purified forms. These assays can also include formats where the added antibodies would prevent T cells from being cross-linked to appropriate antigen presenting cells.

5 Cholera is an ancient, potentially lethal disease caused by the bacterium *Vibrio cholerae* and an effective vaccine for its prevention is still unavailable. Much of the pathogenesis of this disease is caused by the cholera enterotoxin. Ingestion of microgram quantities of cholera toxin can induce severe diarrhea causing loss of tens of liters of fluid.

10

Cholera toxin is a complex of a single catalytic A subunit with a pentameric ring of identical B subunits. Each subunit is inactive on its own. The B subunits bind to specific ganglioside receptors on the surface of intestinal epithelial cells and trigger the entry of the A subunit into the cell. The A subunit
15 ADP-ribosylates a regulatory G protein initiating a cascade of events causing a massive, sustained flow of electrolytes and water into the intestinal lumen resulting in extreme diarrhea.

The B subunit of cholera toxin is an attractive vaccine target for a number
20 of reasons. It is a major target of protective antibodies generated during cholera infection and contains the epitopes for antitoxin neutralizing antibodies. It is nontoxic without the A subunit, is orally effective, and stimulates production of a strong IgA- dominated gut mucosal immune response, which is essential in protection against cholera and cholera toxin. The B subunit is also being
25 investigated for use as an adjuvant in other vaccine preparations, and therefore, evolved toxins may provide general improvements for a variety of different vaccines. The heat-labile enterotoxins (LT) from enterotoxigenic *E. coli* strains are structurally related to cholera toxin and are 75% identical at the DNA sequence level. To obtain optimized recombinant toxin molecules that exhibit

reduced toxicity and increased ability to induce an immune response that is protective against *V. cholerae* and *E. coli*, the genes that encode the related toxins are subjected to stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly.

5

The recombinant toxins are then tested for one or more of a several desirable traits. For example, one can screen for improved cross-reactivity of antibodies raised against the recombinant toxin polypeptides, for lack of toxicity in a cell culture assay, and for ability to induce a protective immune response against the pathogens and/or against the toxins themselves. The experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) clones can be selected by phage display and/or screened by phage ELISA and ELISA assays for the presence of epitopes from the different serotypes. Variant proteins with multiple epitopes can then be purified and used to immunize mice or other test animal. The animal serum is then assayed for antibodies to the different B chain subtypes and variants that elicit a broad cross-reactive response will be evaluated further in a virulent challenge model. The *E. coli* and *V. cholerae* toxins can also act as adjuvants that are capable of enhancing mucosal immunity and oral delivery of vaccines and proteins.

20

Accordingly, one can test the library of recombinant toxins for enhancement of the adjuvant activity

25

Experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens can also be screened for improved expression levels and stability of the B chain pentamer, which may be less stable than when in the presence of the A chain in the hexameric complex. Addition of a heat treatment step or denaturing agents such as salts, urea, and/or

- guanidine hydrochloride can be included prior to ELISA assays to measure yields of correctly folded molecules by appropriate antibodies. It is sometimes desirable to screen for stable monomeric B chain molecules, in an ELISA format, for example, using antibodies that bind monomeric, but not pentameric B chains.
- 5 Additionally, the ability of experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) antigens to elicit neutralizing antibodies in an appropriate animal model such as mouse or monkey can be screened. For example, antibodies that bind to the B chain and prevent its binding to its specific ganglioside receptors on the surface of intestinal epithelial
- 10 cells may prevent disease. Similarly antibodies that bind to the B chain and prevent its pentamerization or block A chain binding may be useful in preventing disease.

- The bacterial antigens that can be improved by stochastic (e.g.
- 15 polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly for use as vaccines also include, but are not limited to, *Helicobacter pylori* antigens CagA and VacA (Blaser (1996) *Aliment. Pharmacol. Ther.* 1: 73-7; Blaser and Crabtree (1996) *Am. J Clin. Pathol.* 106: 565-7; Censini et al. (1996) *Proc. Nat'l. Acad. Sci. USA* 93: 14648-14643).

20

- Other suitable *H. pylori* antigens include, for example, four immunoreactive proteins of 45-65 kDa as reported by Chatha et al. (1997) *Indian J Med. Res.* 105: 170- 175 and the *H. pylori* GroES homologue (HspA) (Kansau et al. (1996) *Mol. Microbiol.* 22: 1013-1023. Other suitable bacterial antigens
- 25 include, but are not limited to, the 43-kDa and the fimbriin (41 kDa) proteins of *P. gingivalis* (Bouts et al. (1996) *Oral Microbiol. Immunol.* 11: 236- 241); pneumococcal surface protein A (Briles et al. (1996) *Ann. NYAcad. Sci.* 797: 118- 126); *Chlamydia psittaci* antigens, 80-90 kDa protein and 110 kDa protein (Buendia et al. (1997) *FEMS Microbiol. Lett.* 150: 113-9); the chlamydial

- exoglycolipid antigen (GLXA) (Whittum-Hudson et al. (1996) *Nature Med.* 2: 1116-1121); *Chlamydia pneumoniae* species-specific antigens in the molecular weight ranges 92-98, 51-55, 43-46 and 31.5-33 kDa and genus-specific antigens in the ranges 12, 26 and 65-70 kDa (Halme et al. (1997) *Scand. J Immunol.* 45: 378-84); *Neisseria gonorrhoeae* (GC) or *Escherichia coli* phase-variable opacity (Opa) proteins (Chen and Gotschlich (1996) *Proc. Nat'l. Acad. Sci. USA* 93: 14851-14856), any of the twelve immunodominant proteins of *Schistosoma mansoni* (ranging in molecular weight from 14 to 208 kDa) as described by Cutts and Wilson (1997) *Parasitology* 114: 245-55; the 17-kDa protein antigen of *Brucella abortus* (De Mot et al. (1996) *Curr. Microbiol.* 33: 26-30); a gene homolog of the 17-kDa protein antigen of the Gram-negative pathogen *Brucella abortus* identified in the nocardioform actinomycete *Rhodococcus* sp. N186/21 (De Mot et al. (1996) *Curr. Microbiol.* 33: 26-30); the staphylococcal enterotoxins (SEs) (Wood et al. (1997) *FEMS Immunol. Med. Microbiol.* 17: 1-10), a 42-kDa *M. hyopneumoniae* NrdF ribonucleotide reductase R2 protein or 15-kDa subunit protein of *M. hyopneumoniae* (Fagan et al. (1997) *Infect. Immun.* 65: 2502-2507), the meningococcal antigen PorA protein (Feavers et al. (1997) *Clin. Diagn. Lab. Immunol.* 3: 444-50); pneumococcal surface protein A (PspA) (McDaniel et al. (1997) *Gene Ther.* 4: 375-377); *F. tularensis* outer membrane protein FopA (Fulop et al. (1996) *FEMS Immunol. Med. Microbiol.* 13: 245-247); the major outer membrane protein within strains of the genus *Actinobacillus* (Hartmann et al. (1996) *Zentralbl. Bakteriol.* 284: 255-262); p60 or listeriolysin (Hly) antigen of *Listeria monocytogenes* (Hess et al. (1996) *Proc. Nat'l. Acad. Sci. USA* 93: 1458-1463); flagellar (G) antigens observed on *Salmonella enteritidis* and *S. pullorum* (Holt and Chabal (1997) *J. Clin. Microbiol.* 35: 1016-1020); *Bacillus anthracis* protective antigen (PA) (Ivins et al. (1995) *Vaccine* 13: 1779-1784); *Echinococcus granulosus* antigen 5 (Jones et al. (1996) *Parasitology* 113: 213-222); the *rol* genes of *Shigella dysenteriae* I and *Escherichia coli* K-12 (Klee et al. (1997) *J. Bacteriol.* 179: 2421 - 2425); cell surface proteins Rib and

- alpha of group B streptococcus (Larsson et al. (1996) *Infect. Immun.* 64: 3518-3523); the 37 kDa secreted polypeptide encoded on the 70 kb virulence plasmid of pathogenic *Yersinia* spp. (Leary et al. (1995) *Contrib. Microbiol. Immunol.* 13: 216-217 and Roggenkamp et al. (1997) *Infect. Immun.* 65: 446- 5 1); the OspA
- 5 (outer surface protein A) of the Lyme disease spirochete *Borrelia burgdorferi* (Li et al. (1997) *Proc. Nat'l. Acad. Sci. USA* 94: 3584-3589, Padilla et al. (1996) *J Infect. Dis.* 174: 739-746, and Wallich et al. (1996) *Infection* 24: 396-397); the *Brucella melitensis* group 3 antigen gene encoding Omp28 (Lindler et al. (1996) *Infect. Immun.* 64: 2490-2499); the PAc antigen of *Streptococcus mutans*
- 10 (Murakami et al. (1997) *Infect. Immun.* 65: 794-797); pneumolysin, *Pneumococcal* neuraminidases, autolysin, hyaluronidase, and the 37 kDa *pneumococcal* surface adhesin A (Paton et al. (1997) *Microb. Drug Resist.* 3: 1 - 10); 29-32, 41-45, 63-71 x 10(3) MW antigens of *Salmonella typhi* (Perez et al. (1996) *Immunology* 89: 262-267); K-antigen as a marker of *Klebsiella*
- 15 *pneumoniae* (Priamukhina and Morozova (1996) *Klin. Lab. Diagn.* 47-9); nocardial antigens of molecular mass approximately 60, 40, and 15-10 kDa (Prokesova et al. (1996) *Int. J Immunopharmacol.* 18: 661- 668); *Staphylococcus aureus* antigen ORF-2 (Rieneck et al. (1997) *Biochim Biophys Acta* 1350: 128-132); GlpQ antigen of *Borrelia hermsii* (Schwan et al. (1996) *J Clin. Microbiol.*
- 20 34: 2483-2492); cholera protective antigen (CPA) (Sciortino (1996) *J. Diarrhoeal Dis. Res.* 14: 16-26); a 190-kDa protein antigen of *Streptococcus mutans* (Senpuku et al. (1996) *Oral Microbiol. Immunol.* 11: 121-128); Anthrax toxin protective antigen (PA) (Sharma et al. (1996) *Protein Expr. Purif.* 7: 33-38); *Clostridium perfringens* antigens and toxoid (Strom et al. (1995) *Br. J.*
- 25 *Rheumatol.* 34: 1095-1096); the SEF14 fimbrial antigen of *Salmonella enteritidis* (Thorns et al. (1996) *Microb. Pathog.* 20: 235-246); the *Yersinia pestis* capsular antigen (F I antigen) (Titball et al. (1997) *Infect. Immun.* 65: 1926- 1930); a 35-kilodalton protein of *Mycobacterium leprae* (Triccas et al. (1996) *Infect. Immun.* 64: 5171-5177); the major outer membrane protein, CD, extracted from *Moraxella*

- (*Branhamella*) *catarrhalis* (Yang et al. (1997) FEMS Immunol. Med. Microbiol. 17: 187-199); pH6 antigen (PsaA protein) of *Yersinia pestis* (Zav'yalov et al. (1996) FEMS Immunol. Med. Microbiol. 14: 53- 57); a major surface glycoprotein, gp63, of *Leishmania major* (Xu and Liew (1994) Vaccine 12: 1534-1536; Xu and Liew (1995) Immunology 84: 173-176); mycobacterial heat shock protein 65, mycobacterial antigen (*Mycobacterium leprae* hsp65) (Lowrie et al. (1994) Vaccine 12: 1537-1540; Ragno et al. (1997) Arthritis Rheum. 40: 277-283; Silva (1995) Braz. J Med. Biol. Res. 28: 843-851); *Mycobacterium tuberculosis* antigen 85 (Ag85) (Huygen et al. (1996) Nat. Med. 2: 893-898); the 45/47 kDa antigen complex (APA) of *Mycobacterium tuberculosis*, *M. bovis* and BCG (Horn et al. (1996) J Immunol. Methods 197: 151-159); the mycobacterial antigen, 65-kDa heat shock protein, hsp65 (Tascon et al. (1996) Nat. Med. 2: 888-892); the mycobacterial antigens MPB64, MPB70, MPB57 and alpha antigen (Yamada et al. (1995) Kekkaku 70: 639-644); the *M. tuberculosis* 38 kDa protein (Vordenneier et al. (1995) Vaccine 13: 1576-1582); the MPT63, MPT64 and MPT- 59 antigens from *Mycobacterium tuberculosis* (Manca et al. (1997) Infect. Immun. 65: 16- 23; Oettinger et al. (1997) Scand. J Immunol. 45: 499-503; Wilcke et al. (1996) Tuber. Lung Dis. 77: 250-256); the 35-kilodalton protein of *Mycobacterium leprae* (Triccas et al. (1996) Infect. Immun. 64: 5171-5177); the ESAT-6 antigen of virulent mycobacteria (Brandt et al. (1996) J Immunol. 157: 3527-3533; Pollock and Andersen (1997) J Infect. Dis. 175: 1251- 1254); *A.~* *mycobacterium tuberculosis* 16-kDa antigen (Hsp16.3) (Chang et al. (1996) J Biol. Chem. 271: 7218-7223); and the 18-kilodalton protein of *Mycobacterium leprae* (Baumgart et al. (1996) Infect. Immun. 64: 2274-2281).

25

2.9.1.2. VIRAL PATHOGENS

The methods of the invention are also useful for obtaining recombinant
5 nucleic acids and polypeptides that have enhanced ability to induce an immune
response against viral pathogens. While the bacterial recombinants described
above are typically administered in polypeptide form, recombinants that confer
viral protection are preferably administered in nucleic acid form, as genetic
vaccines.

10

One illustrative example is the Hantaan virus. Glycoproteins of this virus
typically accumulate at the membranes of the Golgi apparatus of infected cells.
This poor expression of the glycoprotein prevents the development of efficient
genetic vaccines against these viruses. The methods of the invention solve this
15 problem by performing stochastic (e.g. polynucleotide shuffling & interrupted
synthesis) and non-stochastic polynucleotide reassembly on nucleic acids that
encode the glycoproteins and identifying those recombinants that exhibit
enhanced expression in a host cell, and/or for improved immunogenicity when
administered as a genetic vaccine. A convenient screening method for these
20 methods is to express the experimentally generated polynucleotides as fusion
proteins to PIG, which results in display of the polypeptides on the surface of the
host cell (Whitehorn et al. (1995) Biotechnology (N Y) 13:1215-9). Fluorescence-
activated cell sorting is then used to sort and recover those cells that express an
increased amount of the antigenic polypeptide on the cell surface. This
25 preliminary screen can be followed by immunogenicity tests in mammals, such as
mice. Finally, in preferred embodiments, those recombinant nucleic acids are
tested as genetic vaccines for their ability to protect a test animal against
challenge by the virus.

The flaviviruses are another example of a viral pathogen for which the methods of the invention are useful for obtaining a experimentally generated polypeptide or genetic vaccine that is effective against a viral pathogen. The flaviviruses consist of three clusters of antigenically related viruses: Dengue 1-4
5 (62-77% identity), Japanese, St. Louis and Murray Valley encephalitis viruses (75-82% identity), and the tick-borne encephalitis viruses (77- 96% identity). Dengue virus can induce protective antibodies against SLE and Yellow fever (40-50% identity), but few efficient vaccines are available. To obtain genetic vaccines and experimentally generated polypeptides that exhibit enhanced cross-reactivity
10 and immunogenicity, the polynucleotides that encode envelope proteins of related viruses are subjected to stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly. The resulting experimentally generated polynucleotides can be tested, either as genetic vaccines or by using the expressed polypeptides, for ability to induce a broadly reacting
15 neutralizing antibody response. Finally, those clones that are favorable in the preliminary screens can be tested for ability to protect a test animal against viral challenge.

Viral antigens that can be evolved by stochastic (e.g. polynucleotide
20 shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly for improved activity as vaccines include, but are not limited to, influenza A virus N2 neuraminidase (Kilbourne et al. (1995) Vaccine 13: 1799-1803); Dengue virus envelope (E) and premembrane (prM) antigens (Feighny et al. (1994) Am. J Trop. Med. Hyg. 50: 322-328; Putnak et al. (1996) Am. J Trop. Med. Hyg. 55: 504-
25 10); HIV antigens Gag, Pol, Vif and Nef (Vogt et al. (1995) Vaccine 13: 202-208); HIV antigens gp 120 and gp 160 (Achour et al. (1995) Cell. Mol. Biol. 41: 395-400; Hone et al. (1994) Dev. Biol. Stand. 82: 159-162); gp41 epitope of human immunodeficiency virus (Eckhart et al. (1996) J Gen. Virol. 77: 2001- 2008); rotavirus antigen VP4 (Mattion et al. (1995) J Virol. 69: 5132-5137); the rotavirus

- protein VP7 or VP7sc (Emslie et al. (1995) *J Virol.* 69: 1747-1754; Xu et al. (1995) *J Gen. Virol.* 76: 1971-1980); herpes simplex virus (HSV) glycoproteins gB, gC, gD, gE, gG, gH, and gI (Fleck et al. (1994) *Med. Microbiol. Immunol. (Berl)* 183: 87-94 [Mattion, 1995]; Ghiasi et al. (1995) *Invest. Ophthalmol. Vis. Sci.* 36: 1352-1360; McLean et al. (1994) *J Infect. Dis.* 170: 1100-1109);
- 5 immediate-early protein ICP47 of herpes simplex virus- type 1 (HSV-1) (Banks et al. (1994) *Virology* 200: 23 6-245); immediate-early (IE) proteins ICP27, ICPO, and ICP4 of herpes simplex virus (Manickan et al. (1995) *J Virol.* 69: 4711-4716); influenza virus nucleoprotein and hemagglutinin (Deck et al. (1997) *Vaccine* 15: 71- 78; Fu et al. (1997) *J Virol.* 71: 2715-272 1); B 19 parvovirus capsid proteins VP1 (Kawase et al. (1995) *Virology* 211: 359-366) or VP2 (Brown et al. (1994) *Virology* 198: 477- 488); Hepatitis B virus core and e antigen (Schodel et al. (1996) *Intervirology* 39:104-106); hepatitis B surface antigen (Shiau and Murray (1997) *J. Med. Virol.* 51: 159-166); hepatitis B surface antigen fused to the core
- 15 antigen of the virus (Id.); Hepatitis B virus core-preS2 particles (Nemeckova et al. (1996) *Acta Virol.* 40: 273-279); HBV preS2-S protein (Kutinova et al. (1996) *Vaccine* 14: 1045-1052); VZV glycoprotein I (Kutinova et al. (1996) *Vaccine* 14: 1045-1052); rabies virus glycoproteins (Xiang et al. (1994) *Virology* 199: 132-140; Xuan et al. (1995) *Virus Res.* 36: 151-161) or ribonucleocapsid (Hooper et al. (1994) *Proc. Nat'l. Acad. Sci. USA* 91: 10908-10912); human cytomegalovirus (HCMV) glycoprotein B (LTL55) (Britt et al. (1995) *J Infect. Dis.* 171: 18-25); the hepatitis C virus (HCV) nucleocapsid protein in a secreted or a nonsecreted form, or as a fusion protein with the middle (pre-S2 and S) or major (S) surface antigens of hepatitis B virus (HBV) (Inchauspe et al. (1997) *DNA Cell Biol.* 16: 185-195; Major et al. (1995) *J Virol.* 69: 5798-5805); the hepatitis C virus
- 25 antigens: the core protein (pC); E1 (pE1) and E2 (pE2) alone or as fusion proteins (Saito et al. (1997) *Gastroenterology* 112: 1321-1330); the gene encoding respiratory syncytial virus fusion protein (PFP-2) (Falsey and Walsh (1996) *Vaccine* 14: 1214-1218; Piedra et al. (1996) *Pediatr. Infect. Dis. J.* 15: 23-3 1); the

VP6 and VP7 genes of rotaviruses (Choi et al. (1997) *Virology* 232: 129-138; Jin et al. (1996) *Arch. Virol.* 141: 2057-2076); the E1, E2, E3, E4, E5, E6 and E7 proteins of human papillomavirus (Brown et al. (1994) *Virology* 201: 46-54; Dillner et al. (1995) *Cancer Detect. Prev.* 19: 381-393; Krul et al. (1996) *Cancer*
5 *Immunol. Immunother.* 43: 44-48; Nakagawa et al. (1997) *J Infect. Dis.* 175: 927-931); a human T-lymphotropic virus type I gag protein (Porter et al. (1995) *J Med Virol.* 45: 469-474); Epstein-Barr virus (EBV) gp340 (Mackett et al. (1996) *J Med. Virol.* 50: 263-271); the Epstein-Barr virus (EBV) latent membrane protein LMP2 (Lee et al. (1996) *Eur. J Immunol.* 26: 1875-1883); Epstein-Barr virus
10 nuclear antigens 1 and 2 (Chen and Cooper (1996) *J Virol.* 70: 4849-4853; Khanna et al. (1995) *Virology* 214: 633-637); the measles virus nucleoprotein (N) (Fooks et al. (1995) *Virology* 210: 456-465); and cytomegalovirus glycoprotein gB (Marshall et al. (1994) *J Med. Virol.* 43: 77-83) or glycoprotein gH (Rasmussen et al. (1994) *J Infect. Dis.* 170: 673-677).

15

2.9.2. INFLAMMATORY AND AUTOIMMUNE DISEASES

Autoimmune diseases are characterized by immune response that attacks
20 tissues or cells of one's own body, or pathogen-specific immune responses that also are harmful for one's own tissues or cells, or non-specific immune activation which is harmful for one's own tissues or cells. Examples of autoimmune diseases include, but are not limited to, rheumatoid arthritis, SLE, diabetes mellitus, myasthenia gravis, reactive arthritis, ankylosing spondylitis, and multiple
25 sclerosis. These and other inflammatory conditions, including IBD, psoriasis, pancreatitis, and various immunodeficiencies, can be treated using genetic vaccines that include vectors and other components obtained using the methods of the invention (e.g. using antigens that are optimized using the methods of the invention).

These conditions are often characterized by an accumulation of inflammatory cells, such as lymphocytes, macrophages, and neutrophils, at the sites of inflammation. Altered cytokine production levels are often observed, with increased levels of cytokine production. Several autoimmune diseases, including diabetes and rheumatoid arthritis, are linked to certain MHC haplotypes. Other autoimmune-type disorders, such as reactive arthritis, have been shown to be triggered by bacteria such as Yersinia and Shigella, and evidence suggests that several other autoimmune diseases, such as diabetes, multiple sclerosis, rheumatoid arthritis, may also be initiated by viral or bacterial infections in genetically susceptible individuals.

Current strategies of treatment generally include anti-inflammatory drugs, such as NSAID or cyclosporin, and antiproliferative drugs, such as methotrexate. These therapies are non-specific, so a need exists for therapies having greater specificity, and for means to direct the immune responses towards the direction that inhibits the autoimmune process.

The present invention provides several strategies by which these needs can be fulfilled. First, the invention provides methods of obtaining vaccines which exhibit improved delivery of tolerogenic antigens (e.g. methods of obtaining antigens having greater tolerogenicity and/or have improved antigenicity), antigens which have improved antigenicity, genetic vaccine-mediated tolerance, and modulation of the immune response by inclusion of appropriate accessory molecules. In a preferred embodiment, the vaccines (e.g. optimized antigens) prepared according to the invention exhibit improved induction of tolerance by oral delivery.

Oral tolerance is characterized by induction of immunological tolerance after oral administration of large quantities of antigen (Chen et al. (1995) Science 265: 123 7- 1240; Haq et al. (1995) Science 268: 714-716). In animal models, this approach has proven to be a very promising approach to treat autoimmune diseases, and clinical trials are in progress to address the efficacy of this approach in the treatment of human autoimmune diseases, such as rheumatoid arthritis and multiple sclerosis (Chen et al. (1994) Science 265:123 7-40; Whitacre et al. (1996) Clin. Immunol. Immunopathol. 80: S31-9; Hohol et al. (1996) Ann. N.Y Acad Sci. 778:243-50). It has also been suggested that induction of oral tolerance against viruses used in gene therapy might reduce the immunogenicity of gene therapy vectors.

However, the amounts of antigen required for induction of oral tolerance are very high and improved methods for oral delivery of antigenic proteins would significantly improve the efficacy of induction of oral tolerance.

Expression library immunization (Barry et al. (1995) Nature 3 77: 632) is a particularly useful method of screening for optimal antigens for use in genetic vaccines. For example, to identify autoantigens present in Yersinia, Shigella, and the like, one can screen for induction of T cell responses in HLA-B27 positive individuals. Complexes that include epitopes of bacterial antigens and MHC molecules associated with autoimmune diseases, e.g., HLA-B27 in association with Yersinia antigens can be used in the prevention of reactive arthritis and ankylosing spondylitis in HLA-B27 positive individuals.

Treatment of autoimmune and inflammatory conditions can involve not only administration of tolerogenic antigens, but also the use of a combination of cytokines, costimulatory molecules, and the like. Such cocktails are formulated

for induction of a favorable immune response, typically induction of autoantigen-specific tolerance. Cocktails can also include, for example, CD1, which is crucially involved in recognition of self antigens by a subset of T cells (Porcelli (1995) Adv. Immunol. 59: 1). Genetic vaccine vectors and cocktails that skew
5 immune responses towards the T_H2 are often used in treating autoimmune and inflammatory conditions, both with antigen-specific and antigen non-specific vectors.

10 Screening of genetic vaccines and accessory molecules (e.g. and optimized antigens) can be done in animal models which are known to those of skill in the art. Examples of suitable models for various conditions include collagen induced arthritis, the NFS/sld mouse model of human Sjogren's syndrome; a 120 kD organ-specific autoantigen recently identified as an analog of
15 human cytoskeletal protein - fodrin (Haneji et al. (1997) Science 276: 604), the New Zealand Black/White F1 hybrid mouse model of human SLE, NOD mice, a mouse model of human diabetes mellitus, fas/fas ligand mutant mice, which spontaneously develop autoimmune and lymphoproliferative disorders (Watanabe-Fukunaga et al. (1992) Nature 356: 314), and experimental
20 autoimmune encephalomyelitis (EAE), in which myelin basic protein induces a disease that resembles human multiple sclerosis.

Autoantigens (that can be experimentally evolved (e.g. by polynucleotide
25 reassembly &/or polynucleotide site-saturation mutagenesis) according to the methods of the invention) that are useful in genetic vaccines for treating multiple sclerosis include, but are not limited to, myelin basic protein (Stinissen et al. (1996) J Neurosci. Res. 45: 500-511) or a fusion protein of myelin basic protein and proteolipid protein in multiple sclerosis (Elliott et al. (1996) J Clin. Invest. 98:

1602-1612), proteolipid protein (PLP) (Rosener et al. (1997) *J Neuroimmunol.* 75: 28-34), 2',3'-cyclic nucleotide 3'-phosphodiesterase (CNPase) (Rosener et al. (1997) *J Neuroimmunol.* 75: 28-34), the Epstein Barr virus nuclear antigen-1 (EBNA-1) in multiple sclerosis (Vaughan et al. (1996) *J Neuroimmunol.* 69: 95-102), HSP70 in multiple sclerosis (Salvetti et al. (1996) *J Neuroimmunol.* 65: 143-53; Feldmann et al. (1996) *Cell* 85: 307).

Target antigens that, after reassembly (optionally in combination with other directed evolution methods described herein) according to the methods of the invention, can be used to treat scleroderma, systemic sclerosis, and systemic lupus erythematosus include, for example, (-2-GPI, 50 kDa glycoprotein (Blank et al. (1994) *J Autoimmun.* 7: 441-455), Ku (p70/p80) autoantigen, or its 80-kd subunit protein (Hong et al. (1994) *Invest. Ophthalmol. Vis. Sci.* 35: 4023-4030; Wang et al. (1994) *J Cell Sci.* 107: 3223-3233), the nuclear autoantigens La (SS-B) and Ro (SS-A) (Huang et al. (1997) *J Clin. Immunol.* 17: 212-219; Igarashi et al. (1995) *Autoimmunity* 22: 33-42; Keech et al. (1996) *Clin. Exp. Immunol.* 104: 255-263; Manoussakis et al. (1995) *J Autoimmun.* 8: 959-969; Topfer et al. (1995) *Proc. Nat'l. Acad. Sci. USA* 92: 875-879), proteasome (-type subunit C9 (Feist et al. (1996) *J Exp. Med.* 184: 1313-1318), Scleroderma antigens Rpp 30, Rpp 38 or Scl-70 (Eder et al. (1997) *Proc. Nat'l. Acad. Sci. USA* 94: 1101-1106; Hietarinta et al. (1994) *Br. J Rheumatol.* 33: 323-326), the centrosome autoantigen PCM-1 (Bao et al. (1995) *Autoimmunity* 22: 219-228), polymyositis-scleroderma autoantigen (PM-Scl) (Kho et al. (1997) *J Biol. Chem.* 272: 13426-13431), scleroderma (and other systemic autoimmune disease) autoantigen CENP-A (Muro et al. (1996) *Clin. Immunol. Immunopathol.* 78: 86-89), U5, a small nuclear ribonucleoprotein (snRNP) (Okano et al. (1996) *Clin. Immunol. Immunopathol.* 81: 41-47), the 100-kd protein of PM-Scl autoantigen (Ge et al. (1996) *Arthritis Rheum.* 39: 1588-1595), the nucleolar U3- and Th(7-2) ribonucleoproteins (Verheijen et al. (1994) *J. Immunol. Methods* 169: 173-182),

the ribosomal protein L7 (Neu et al. (1995) Clin. Exp. Immunol. 100: 198-204), hPop 1 (Lygerou et al. (1996) EMBO J. 15: 5936-5948), and a 36-kd protein from nuclear matrix antigen (Deng et al. (1996) Arthritis Rheum. 39: 1300-1307).

- 5 Hepatic autoimmune disorders can also be treated using improved recombinant antigens that are prepared according to the methods described herein. Among the antigens that are useful in such treatments are the cytochromes P450 and UDP- glucuronosyl-transferases (Obermayer-Straub and Manns (1996) Baillieres Clin. Gastroenterol. 10: 501-532), the cytochromes P450 2C9 and P450
- 10 1A2 (Bourdi et al. (1996) Chem. Res. Toxicol. 9: 1159-1166; Clemente et al. (1997) J Clin. Endocrinol. Metab. 82: 1353-1361), LC-1 antigen (Klein et al. (1996) J Pediatr. Gastroenterol. Nutr. 23: 461-465), and a 230-kDa Golgi-associated protein (Funaki et al. (1996) Cell Struct. Funct. 21: 63-72).
- 15 For treatment of autoimmune disorders of the skin, useful antigens include, but are not limited to, the 450 kD human epidermal autoantigen (Fujiwara et al. (1996) J Invest. Dermatol. 106: 1125-1130), the 230 kD and 180 kD bullous pemphigoid antigens (Hashimoto (1995) Keio J Med. 44: 115 -123; Murakami et al. (1996) J Dermatol. Sci. 13: 112-117), pemphigus foliaceus
- 20 antigen (desmoglein 1), pemphigus vulgaris antigen (desmoglein 3), BPAg2, BPAg1, and type VII collagen (Batteux et al. (1997) J Clin. Immunol. 17: 228-233; Hashimoto et al. (1996) J Dermatol. Sci. 12: 10- 17), a 168-kDa mucosal antigen in a subset of patients with cicatricial pemphigoid (Ghohestani et al. (1996) J Invest. Dermatol. 107: 136-139), and a 218-kd nuclear protein (218-kd
- 25 Mi-2) (Seelig et al. (1995) Arthritis Rheum. 38: 1389-1399).

The methods of the invention are also useful for obtaining improved antigens for treating insulin dependent diabetes mellitus, using one or more of antigens which include, but are not limited to, insulin, proinsulin, GAD65 and

- GAD67, heat-shock protein 65 (hsp65), and islet-cell antigen 69 (ICA69) (French et al. (1997) *Diabetes* 46: 34-39; Roep (1996) *Diabetes* 45: 1147-1156; Schloot et al. (1997) *Diabetologia* 40: 332-338), viral proteins homologous to GAD65 (Jones and Crosby (1996) *Diabetologia* 39: 1318-1324), islet cell antigen-related protein-tyrosine phosphatase (PTP) (Cui et al. (1996) *J Biol. Chem.* 271: 24817-24823), GM2-1 ganglioside (Cavallo et al. (1996) *J Endocrinol.* 150: 113-120; Dotta et al. (1996) *Diabetes* 45: 1193-1196), glutamic acid decarboxylase (GAD) (Nepom (1995) *Curr. Opin. Immunol.* 7: 825-830; Panina-Bordignon et al. (1995) *J Exp. Med.* 181: 1923-1927), an islet cell antigen (ICA69) (Karges et al. (1997) *Biochim. Biophys. Acta* 1360: 97-101; Roep et al. (1996) *Eur. J Immunol.* 26: 1285-1289), Tep69, the single T cell epitope recognized by T cells from diabetes patients (Karges et al. (1997) *Biochim. Biophys. Acta* 1360: 97-101), ICA 512, an autoantigen of type I diabetes (Solimena et al. (1996) *EMBOJ.* 15: 2102-2114), an islet-cell protein tyrosine phosphatase and the 37- kDa autoantigen derived from it in type I diabetes (including IA-2, IA-2) (La Gasse et al. (1997) *Mol. Med.* 3: 163-173), the 64 kDa protein from In- 111 cells or human thyroid follicular cells that is immunoprecipitated with sera from patients with islet cell surface antibodies (ICSA) (Igawa et al. (1996) *Endocr. J.* 43: 299-306), phogrin, a homologue of the human transmembrane protein tyrosine phosphatase, an autoantigen of type I diabetes (Kawasaki et al. (1996) *Biochem. Biophys. Res. Commun.* 227: 440-447), the 40 kDa and 37 kDa tryptic fragments and their precursors IA-2 and IA-2 in IDDM (Lampasona et al. (1996) *J Immunol.* 157: 2707-2711; Notkins et al. (1996) *J Autoimmun.* 9: 677-682), insulin or a cholera toxoid- insulin conjugate (Bergerot et al. (1997) *Proc. Nat'l. Acad. Sci. USA* 94: 4610-4614), carboxypeptidase H, the human homologue of gp330, which is a renal epithelial glycoprotein involved in inducing Heymann nephritis in rats, and the 38- kD islet mitochondrial autoantigen (Arden et al. (1996) *J Clin. Invest.* 97: 551- 561.

Rheumatoid arthritis is another condition that is treatable using optimized antigens prepared according to the present invention. Useful antigens for rheumatoid arthritis treatment include, but are not limited to, the 45 kDa DEK nuclear antigen, in particular onset juvenile rheumatoid arthritis and iridocyclitis (Murray et al. (1997) *J Rheumatol.* 24: 560- 567), human cartilage glycoprotein-39, an autoantigen in rheumatoid arthritis (Verheijden et al. (1997) *Arthritis Rheum.* 40: 1115-1125), a 68k autoantigen in rheumatoid arthritis (Blass et al. (1997) *Ann. Rheum. Dis.* 56: 317-322), collagen (Rosloniec et al. (1995) *J Immunol.* 155: 4504-4511), collagen type II (Cook et al. (1996) *Arthritis Rheum.* 39: 1720-1727; Trentham (1996) *Ann. N. Y. Acad. Sci.* 778: 306-314), cartilage link protein (Guerassimov et al. (1997) *J Rheumatol.* 24: 95 9-964), ezrin, radixin and moesin, which are auto-immune antigens in rheumatoid arthritis (Wagatsuma et al. (1996) *Mol. Immunol.* 33: 1171-1176), and mycobacterial heat shock protein 65 (Ragno et al. (1997) *Arthritis Rheum.* 40: 277-283).

15

Also among the conditions for which one can obtain an improved antigen suitable for treatment are autoimmune thyroid disorders. Antigens that are useful for these applications include, for example, thyroid peroxidase and the thyroid stimulating hormone receptor (Tandon and Weetman (1994) *J R. Coll. Physicians Lond.* 28: 10- 18), thyroid peroxidase from human Graves' thyroid tissue (Gardas et al. (1997) *Biochem. Biophys. Res. Commun.* 234: 366-370; Zimmer et al. (1997) *Histochem. Cell. Biol.* 107: 115-120), a 64-kDa antigen associated with thyroid-associated ophthalmopathy (Zhang et al. (1996) *Clin. Immunol. Immunopathol.* 80: 23 6-244), the human TSH receptor (Nicholson et al. (1996) *J Mol. Endocrinol.* 16: 159-170), and the 64 kDa protein from In- 111 cells or human thyroid follicular cells that is immunoprecipitated with sera from patients with islet cell surface antibodies (ICSA) (Igawa et al. (1996) *Endocr. J.* 43: 299-306).

25

Other conditions and associated antigens include, but are not limited to, Sjogren's syndrome (-fodrin; Haneji et al. (1997) Science 276: 604-607), myasthenia gravis (the human M2 acetylcholine receptor or fragments thereof, specifically the second extracellular loop of the human M2 acetylcholine receptor; 5 Fu et al. (1996) Clin. Immunol. Immunopathol. 78: 203-207), vitiligo (tyrosinase; Fishman et al. (1997) Cancer 79: 1461 - 1464), a 450 kD human epidermal autoantigen recognized by serum from individual with blistering skin disease, and ulcerative colitis (chromosomal proteins HMG1 and HMG2; Sobajima et al. (199 7) Clin. Exp. Immunol. 107: 135 -140).

10

2.9.3. ALLERGY AND ASTHMA

The invention also provides methods of obtaining reagents that are useful 15 for treating allergy. In one embodiment, the methods involve making a library of experimentally generated polynucleotides that encode an allergen, and screening the library to identify those experimentally generated polynucleotides that exhibit improved properties when used as immunotherapeutic reagents for treating allergy. For example, specific immunotherapy of allergy using natural antigens 20 carries a risk of inducing anaphylaxis, which can be initiated by cross-linking of high-affinity IgE receptors on mast cells. Therefore, allergens that are not recognized by pre-existing IgE are desirable. The methods of the invention provide methods by which one can obtain such allergen variants. Another improved property of interest is induction of broader immune responses, 25 increased safety and efficacy.

Genetic vaccine vectors and other reagents obtained using the methods of the invention can be used to treat allergies and asthma. Allergic immune responses are results of complex interactions between B cells, T cells,

professional antigen- presenting cells (APC), eosinophils and mast cells. These cells take part in allergic immune responses both as modulators of the immune responses and are also involved in producing factors directly involved in initiation and maintenance of allergic responses.

5

Synthesis of polyclonal and allergen-specific IgE requires multiple interactions between B cells, T cells and professional antigen- presenting cells (APC).

10

Activation of naive, unprimed B cells is initiated when specific B cells recognize the allergen by cell surface immunoglobulin (sIg). However, costimulatory molecules expressed by activated T cells in both soluble and membrane-bound forms are necessary for differentiation of B cells into IgE-secreting plasma cells. Activation of T helper cells requires recognition of an antigenic peptide in the context of MHC class II molecules on the plasma membrane of APC, such as monocytes, dendritic cells, Langerhans cells or primed B cells. Professional APC can efficiently capture the antigen and the peptide-MHC class II complexes are formed in a post-Golgi, proteolytic intracellular compartment and subsequently exported to the plasma membrane, where they are recognized by T cell receptor (TCR) (Monaco (1995) J Leuk. Biol. 57: 543-547). In addition, activated B cells express CD80 (B7-1) and CD86 (B7-2, B70), which are the counter receptors for CD28 and which provide a costimulatory signal for T cell activation resulting in T cell proliferation and cytokine synthesis (Bluestone (1995) Immunity 2: 555-559). Since allergen-specific T cells from atopic individuals generally belong to the T_H2 cell subset, activation of these cells also leads to production of IL-4 and IL-13, which, together with membrane- bound costimulatory molecules expressed by activated T helper cells, direct B cell differentiation into IgE-secreting plasma cells (de

15

20

25

Vries and Punnonen, In Cytokine Regulation of Humoral Immunity: Basic and Clinical Aspects, Ed. CM Snapper, John Wiley & Sons Ltd, West Sussex, UK, p. 195-215, 1996).

5 Mast cells and eosinophils are key cells in inducing allergic symptoms in target organs. Recognition of specific antigen by IgE bound to high- affinity IgE receptors on mast cells, basophils or eosinophils results in crosslinking of the receptors leading to degranulation of the cells and rapid release of mediator molecules, such as histamine, prostaglandins and leukotrienes, causing allergic
10 symptoms.

 Immunotherapy of allergic diseases currently includes hyposensibilization treatments using increasing doses of allergen injected to the patient. These treatments result skewing of immune responses towards T_H1 phenotype and
15 increase the ratio of IgG/IgE antibodies specific for allergens. Because these patients have circulating IgE antibodies specific for the allergens, these treatments include significant risk of anaphylactic reactions.

 In these reactions, free circulating allergen is recognized by IgE molecules
20 bound to high-affinity IgE receptors on mast cells and eosinophils. Recognition of the allergen results in crosslinking of the receptors leading to release of mediators, such as histamine, prostaglandins, and leukotrienes, which cause the allergic symptoms, and occasionally anaphylactic reactions. Other problems associated with hyposensibilization include low efficacy and difficulties in producing
25 allergen extracts reproducibly.

 Genetic vaccines provide a means of circumventing the problems that have limited the usefulness of previously known hyposensibilization treatments. For example, by expressing antigens on the surface of cells, such as muscle cells,

the risk of anaphylactic reactions is significantly reduced. This can be achieved by using genetic vaccine vectors that encode transmembrane forms of allergens. The allergens can also be modified in such a way that they are efficiently expressed in transmembrane forms, further reducing the risk of anaphylactic reactions. Another
5 advantage provided by the use of genetic vaccines for hyposensibilization is that the genetic vaccines can include cytokines and accessory molecules which further direct the immune responses towards the T_H1 phenotype, thus reducing the amount of IgE antibodies produced and increasing the efficacy of the treatments. Vectors can also be evolved to induce primarily IgG and IgM responses, with little
10 or no IgE response.

Furthermore, stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly can be used to generate allergens that are not recognized by the specific IgE antibodies preexisting *in vivo*,
15 yet are capable of inducing efficient activation of allergen-specific T cells. For example, using phage display selection, one can express experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) allergens on phage, and only those that are not recognized by specific IgE antibodies are selected. These are further screened for their capacity
20 to induce activation of specific T cells. An efficient T cell response is an indication that the T cell epitopes are functionally intact, although the B cell epitopes were altered, as indicated by lack of binding of specific antibodies.

In these methods, polynucleotides encoding known allergens, or homologs
25 or fragments thereof (e.g., immunogenic peptides) are inserted into DNA vaccine vectors and used to immunize allergic and asthmatic individuals. Alternatively, the experimentally evolved (e.g. by polynucleotide reassembly &/or polynucleotide site-saturation mutagenesis) allergens are expressed in manufacturing cells, such as *E. coli* or yeast cells, and subsequently purified and

used to treat the patients or prevent allergic disease. stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly can be used to obtain antigens that activate T cells but cannot induce anaphylactic reactions. For example, a library of experimentally
5 generated polynucleotides that encode allergen variants can be expressed in cells, such as antigen presenting cells, which are then contacted with PBMC or T cell clones from atopic patients. Those library members that efficiently activate T_H cells from the atopic patients can be identified by assaying for T cell proliferation, or by cytokine synthesis (e.g., synthesis of IL-2, IL-4, IFN- γ). Those recombinant
10 allergen variants that are positive in the *in vitro* tests can then be subjected to *in vivo* testing.

Examples of allergies that can be treated include, but are not limited to,
15 allergies against house dust mite, grass pollen, birch pollen, ragweed pollen, hazel pollen, cockroach, rice, olive tree pollen, fmgj, mustard, bee venom.

Antigens of interest include those of animals, including the mite (e.g., Dermatophagoides pteronyssinus, Dermatophagoides farinae, Blomia tropicalis),
20 such as the allergens der p1 (Scobie et al. (1994) Biochem. Soc. Trans. 22: 448S; Yssel et al. (1992) J Immunol. 148: 738-745), der p2 (Chua et al. (1996) Clin. Exp. Allergy 26: 829-837), der p3 (Smith and Thomas (1996) Clin. Exp. Allergy 26: 571-579), der p5, der p V (Lin et al. (1994) J Allergy Clin. Immunol. 94: 989-996), der p6 (Bennett and Thomas (1996) Clin. Exp. Allergy 26: 1150-1154), der
25 p7 (Shen et al. (1995) Clin. Exp. Allergy 25: 416-422), der f2 (Yuuki et al. (1997) Int. Arch. Allergy Immunol. 112: 44-48), der f3 (Nishiyama et al. (1995) FEBS Lett. 377: 62-66), der f7 (Shen et al. (1995) Clin. Exp. Allergy 25: 1000-1006); Mag 3 (Fujikawa et al. (1996) Mol. Immunol. 33: 311-319). Also of interest as antigens are the house dust mite allergens Tyr p2 (Eriksson et al. (1998)

- Eur. J Biochem. 251: 443-447), Lep d 1 (Schmidt et al. (1995) FEBS Lett. 370: 11-14), and glutathione S-transferase (O'Neill et al. (1995) Immunol Lett. 48: 103-107); the 25,589 Da, 219 amino acid polypeptide with homology with glutathione S-transferases (O'Neill et al. (1994) Biochim. Biophys. Acta. 1219: 521-528); Blo t 5 (Arruda et al. (1995) Int. Arch. Allergy Immunol. 107: 456-457); bee venom phospholipase A2 (Carballido et al. (1994) J Allergy Clin. Immunol. 93: 758-767; Jutel et al. (1995) J Immunol. 154: 4187-4194); bovine dermal/dander antigens BDA 11 (Rautiainen et al. (1995) J. Invest. Dermatol. 105: 660-663) and BDA20 (Mantylä et al. (1996) J Allergy Clin. Immunol. 97: 1297-1303); the major horse allergen Equ c1 (Gregoire et al. (1996) J Biol. Chem. 271: 32951-32959); Jumper ant M. pilosula allergen Myr p 1 and its homologous allergenic polypeptides Myr p2 (Donovan et al. (1996) Biochem. Mol. Biol. Int. 39: 877-885); 1-13, 14, 16 kD allergens of the mite Blomia tropicalis (Caraballo et al. (1996) J Allergy Clin. Immunol. 98: 573-579); the cockroach allergens Bla g 2 (Helm et al. (1996) J Allergy Clin. Immunol. 98: 172-80) and Bla g 2 (Arruda et al. (1995) J Biol. Chem. 270: 19563-19568); the cockroach Cr-PI allergens (Wu et al. (1996) J Biol. Chem. 271: 17937-17943); fire ant venom allergen, Sol i 2 (Schmidt et al. (1996) J Allergy Clin. Immunol. 98: 82-88); the insect Chironomus thumini major allergen Chi t 1-9 (Kipp et al. (1996) Int. Arch. Allergy Immunol. 110: 348-353); dog allergen Can f 1 or cat allergen Fel d 1 (Ingram et al. (1995) J Allergy Clin. Immunol. 96: 449-456); albumin, derived, for example, from horse, dog or cat (Goubran Botros et al. (1996) Immunology 88: 340-347); deer allergens with the molecular mass of 22 kD, 25 kD or 60 kD (Spitzauer et al. (1997) Clin. Exp. Allergy 27: 196-200); and the 20 kD major allergen of cow (Ylonen et al. (1994) J Allergy Clin. Immunol. 93: 851-858).
- Pollen and grass allergens are also useful in vaccines, particularly after optimization of the antigen by the methods of the invention. Such allergens include, for example, Hor v9 (Astwood and Hill (1996) Gene 182: 53-62, Lig v 1 (Batanero et al. (1996) Clin. Exp. Allergy 26: 1401-1410); Lol p 1 (Muller et al.

- (1996) *Int. Arch. Allergy Immunol.* 109: 352-355), Lol p II (Tamborini et al. (1995) *Mol. Immunol.* 32: 505- 513), Lol pVA, Lol pVB (Ong et al. (1995) *Mol. Immunol.* 32: 295-302), Lol p 9 (Blaher et al. (1996) *J Allergy Clin. Immunol.* 98: 124-132); Par J I (Costa et al. (1994) *FEBS Lett.* 341: 182-186; Sallusto et al. 5 (1996) *J Allergy Clin. Immunol.* 97: 627-637), Par j 2.0101 (Duro et al. (1996) *FEBS Lett.* 399: 295-298); Bet v1 (Faber et al. (1996) *J Biol. Chem.* 271: 19243-19250), Bet v2 (Rihs et al. (1994) *Int. Arch. Allergy Immunol.* 105: 190-194); Dac g3 (Guerin-Marchand et al. (1996) *Mol. Immunol.* 33: 797-806); Phl p 1 (Petersen et al. (1995) *J Allergy Clin. Immunol.* 95: 987-994), Phl p 5 (Muller et 10 al. (1996) *Int. Arch. Allergy Immunol.* 109: 352- 355), Phl p 6 (Petersen et al. (1995) *Int. Arch. Allergy Immunol.* 108: 55-59); Cry j I (Sone et al. (1994) *Biochem. Biophys. Res. Commun.* 199: 619-625), Cry j II (Namba et al. (1994) *FEBS Lett.* 353: 124-128); Cor a 1 (Schenk et al. (1994) *Eur. J Biochem.* 224: 717-722); cyn d 1 (Smith et al. (1996) *J Allergy Clin. Immunol.* 98: 331-343), cyn 15 d 7 (Suphioglu et al. (1997) *FEBS Lett.* 402: 167-172); Pha a 1 and isoforms of Pha a 5 (Suphioglu and Singh (1995) *Clin. Exp. Allergy* 25: 853-865); Cha o 1 (Suzuki et al. (1996) *Mol. Immunol.* 33: 451-460); profilin derived, e.g, from timothy grass or birch pollen (Valenta et al. (1994) *Biochem. Biophys. Res. Commun.* 199:106-118); P0149(Wuet al. (1996) *Plant Mol.Biol.* 32: 1037-1042); 20 Ory s1 (Xuet al. (1995) *Gene* 164:255-259); and Amb a V and Amb t5 (Kim et al. (1996) *Mol. Immunol.* 33: 873-880; Zhu et al. (1995) *J Immunol.* 155: 5064-5073).

Vaccines against food allergens can also be developed using the methods 25 of the invention. Suitable antigens for reassembly (optionally in combination with other directed evolution methods described herein) include, for example, profilin (Rihs et al. (1994) *Int. Arch. Allergy Immunol.* 105: 190-194); rice allergenic cDNAs belonging to the alpha-amylase/trypsin inhibitor gene family (Alvarez et al. (1995) *Biochim Biophys Acta* 1251: 201-204); the main olive allergen, Ole e I

(Lombardero et al. (1994) Clin Exp Allergy 24: 765-770); Sin a 1, the major allergen from mustard (Gonzalez De La Pena et al. (1996) Eur J Biochem. 237: 827-832); parvalbumin, the major allergen of salmon (Lindstrom et al. (1996) Scand. J Immunol. 44: 335-344); apple allergens, such as the major allergen Mal d 1 (Vanek-Krebitz et al. (1995) Biochem. Biophys. Res. Commun. 214: 538-551); and peanut allergens, such as Ara h I (Burks et al. (1995) J Clin. Invest. 96: 1715- 1721).

The methods of the invention can also be used to develop recombinant antigens that are effective against allergies to fungi. Fungal allergens useful in these vaccines include, but are not limited to, the allergen, Cla h III, of *Cladosporium herbarum* (Zhang et al. (1995) J Immunol. 154: 710-717); the allergen Psi c 2, a fungal cyclophilin, from the basidiomycete *Psilocybe cubensis* (Homer et al. (1995) Int. Arch. Allergy Immunol. 107: 298-300); hsp 70 cloned from a cDNA library of *Cladosporium herbarum* (Zhang et al. (1996) Clin Exp Allergy 26: 88-95); the 68 kD allergen of *Penicillium notatum* (Shen et al. (1995) Clin. Exp. Allergy 26: 350-356); aldehyde dehydrogenase (ALDH) (Achatz et al. (1995) Mol Immunol. 32: 213-227); enolase (Achatz et al. (1995) Mol. Immunol. 32: 213- 227); YCP4 (Id.); acidic ribosomal protein P2 (Id.).

20

Other allergens that can be used in the methods of the invention include latex allergens, such as a major allergen (Hev b 5) from natural rubber latex (Akasawa et al. (1996) J Biol. Chem. 271: 25389-25393; Slater et al. (1996) J Biol. Chem. 271: 25394- 25399).

25

The invention also provides a solution to another shortcoming of genetic vaccination as a treatment for allergy and asthma. While genetic vaccination primarily induces CD8⁺ T cell responses, induction of allergen-specific IgE responses is dependent on CD4⁺ T cells and their help to B cells. T_H2-type cells

are particularly efficient in inducing IgE synthesis because they secrete high levels of IL-4, IL-5 and IL-13, which direct Ig isotype switching to IgE synthesis. IL-5 also induces eosinophilia. The methods of the invention can be used to develop genetic vaccines that efficiently induce CD4⁺ T cell responses, and direct
5 differentiation of these cells towards the T_H1 phenotype.

The invention also provides methods by which the level of antigen release by a genetic vaccine vector is regulated. Regulation of the antigen dose is crucial at the onset of hyposensibilization for safety reasons. Low antigen levels are
10 preferably used at first, with the antigen level increasing once evidence has been obtained that the antigen does not induce adverse effects in the individual. The stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly methods of the invention allow generation of genetic vaccine vectors that induce expression of different (high and low)
15 levels of antigen. For example, two or more different evolved promoters can be used for antigen expression. Alternatively, the antigen gene itself can be evolved for different levels of expression by, for example, altering codon usage. Vectors that induce different levels of antigen expression can be screened by use of specific monoclonal antibodies, and cell sorting (e.g, FACS).

20

2.9.4. CANCER

Immunotherapy has great promise for the treatment of cancer and
25 prevention of metastasis. By inducing an immune response against cancerous cells, the body's immune system can be enlisted to reduce or eliminate cancer. (e.g. using the improved antigens obtained using the methods of the invention). Genetic vaccines prepared using the methods of the invention, as well as

accessory molecules described herein, provide cancer immunotherapies of increased effectiveness compared to those that are presently available.

One approach to cancer immunotherapy is vaccination using genetic vaccines that include or encode antigens that are specific for tumor cells or by injecting the patients with purified recombinant cancer antigens. The methods of the invention can be used for (obtaining antigens that exhibit an) enhancement of immune responses against known tumor-specific antigens, and also to search for novel protective antigenic sequences. Genetic vaccines that exhibit optimized antigen expression, processing, and presentation can be obtained as described herein. The methods of the invention are also suitable for obtaining optimized cytokines, costimulatory molecules, and other accessory molecules that are effective in induction of an antitumor immune response, as well as for obtaining genetic vaccines and cocktails that include these and other components present in optimal combinations. The approach used for each particular cancer can vary. For treatment of hormone-sensitive cancers (for example, breast cancer and prostate cancer), methods of the invention can be used to obtain optimized hormone antagonists. For highly immunogenic tumors, including melanoma, one can screen for genetic vaccine vectors (recombinant antigens) that optimally boost the immune response against the tumor.

Breast cancer, in contrast, is of relatively low immunogenicity and exhibits slow progression, so individual treatments can be designed for each patient. Prevention of metastasis is also a goal in design of genetic vaccines.

25

Among the tumor-specific antigens that can be used in the antigen reassembly (optionally in combination with other directed evolution methods described herein) methods of the invention are: bullous pemphigoid antigen 2, prostate mucin antigen (PMA) (Beckett and Wright (1995) Int. J Cancer 62: 703-

- 710), tumor associated Thomsen- Friedenreich antigen (Dahlenborg et al. (1997) Int. J Cancer 70: 63-71), prostate-specific antigen (PSA) (Dannull and Belldgrun (1997) Br. J Urol. 1: 97-103), luminal epithelial antigen (LEA. 135) of breast carcinoma and bladder transitional cell carcinoma (TCC) (Jones et al. (1997) 5 Anticancer Res. 17: 685-687), cancer-associated serum antigen (CASA) and cancer antigen 125 (CA 125) (Kierkegaard et al. (1995) Gynecol. Oncol. 59: 251-254), the epithelial glycoprotein 40 (EGP40) (Kievit et al. (1997) Int. J Cancer 71: 237-245), squamous cell carcinoma antigen (SCC) (Lozza et al. (1997) Anticancer Res. 17: 525-529), cathepsin E (Mota et al. (1997) Ant. J Pathol. 150: 10 1223-1229), tyrosinase in melanoma (Fishman et al. (1997) Cancer 79: 1461-1464), cell nuclear antigen (PCNA) of cerebral cavernomas (Notelet et al. (1997) Surg. Neurol. 47: 364-370), DF3/MUC1 breast cancer antigen (Apostolopoulos et al. (1996) Immunol. Cell. Biol. 74: 45 7-464; Pandey et al. (1995) Cancer Res. 5 5: 4000-4003), carcinoembryonic antigen (Paone et al. (1996) J Cancer Res. Clin. 15 Oncol. 122: 499-503; Schlom et al. (1996) Breast Cancer Res. Treat. 38: 27-39), tumor-associated antigen CA 19-9 (Tolliver and O'Brien (1997) South Med. J. 90: 89-90; Tsuruta et al. (1997) Urol. Int. 5 8: 20-24), human melanoma antigens MART- I /Melan-A27- and gp100 (Kawakami and Rosenberg (1997) Int. Rev. Immunol. 14: 173-192; Zajac et al. (1997) Int. J Cancer 71: 491-496), the T and 20 Tn pancarcinoma (CA) glycopeptide epitopes (Springer (1995) Crit. Rev. Oncog. 6: 57-85), a 35 kD tumor-associated autoantigen in papillary thyroid carcinoma (Lucas et al. (1996) Anticancer Res. 16: 2493 -2496), KH- I adenocarcinoma antigen (Deshpande and Danishefsky (1997) Nature 387: 164-166), the A60 mycobacterial antigen (Maes et al. (1996) J Cancer Res. Clin. Oncol. 122: 296- 25 300), heat shock proteins (HSPs) (Blachere and Srivastava (1995) Semin. Cancer Biol. 6: 349-355), and MAGE, tyrosinase, melan-A and gp75 and mutant oncogene products (e.g., p53, ras, and HER-2/neu (Bueler and Mulligan (1996) Mol. Med. 2: 545-555; Lewis and Houghton (1995) Semin. Cancer Biol. 6: 321-327; Theobald et al. (1995) Proc. Nat'l. Acad. Sci. USA 92: 11993-11997).

2.9.5. PARASITES

5 Antigens from parasites can also be optimized by the methods of the invention. These include, but are not limited to, the schistosome gut- associated antigens CAA (circulating anodic antigen) and CCA (circulating cathodic antigen) in *Schistosoma mansoni*, *S. haematobium* or *S. japonicum* (Deelder et al. (1996) *Parasitology* 112: 21-35); a multiple antigen peptide (MAP) composed of
10 two distinct protective antigens derived from the parasite *Schistosoma mansoni* (Ferru et al. (1997) *Parasite Immunol.* 19: 1 -11); *Leishmania* parasite surface molecules (Lezama-Davila (1997) *Arch. Med Res.* 28: 47-53); third-stage larval (L3) antigens of *L. loa* (Akue et al. (1997) *J Infect. Dis.* 175: 158-63); the genes, Tams 1-1 and Tams 1-2, encoding the 30-and 32-kDa major merozoite surface
15 antigens of *Theileria annulata* (Ta) (d'Oliveira et al. (1996) *Gene* 172: 33-39); *Plasmodium falciparum* merozoite surface antigen 1 or 2 (al-Yaman et al. (1995) *Trans. R. Soc. Trop. Med. Hyg.* 89: 555-559; Beck et al. (1997) *J Infect. Dis.* 175: 921-926; Rzepczyk et al. (1997) *Infect. Immun.* 65: 1098-1100); circumsporozoite (CS) protein- based B-epitopes from *Plasmodium berghei*,
20 (PPPPNPND)₂ and *Plasmodium yoelii*, (QGPGAP)₃QG, along with a *P. berghei* T-helper epitope KQIRDSITEEWS (Reed et al. (1997) *Vaccine* 15: 482-488); NYVAC-Pf7 encoded *Plasmodium falciparum* antigens derived from the sporozoite (circumsporozoite protein and sporozoite surface protein 2), liver (liver stage antigen 1), blood (merozoite surface protein 1, serine repeat antigen, and
25 apical membrane antigen 1), and sexual (25-kDa sexual-stage antigen) stages of the parasite life cycle were inserted into a single NYVAC genome to generate NYVAC-Pf7 (Tine et al. (1996) *Infect. Immun.* 64: 3833-3844); *Plasmodium falciparum* antigen Pfs230 (Williamson et al. (1996) *Mol. Biochem. Parasitol.* 78: 161-169); *Plasmodium falciparum* apical membrane antigen (AMA- 1) (Lal et al.

- (1996) *Infect. Immun.* 64: 1054-1059); *Plasmodium falciparum* proteins Pfs28 and Pfs25 (Duffy and Kaslow (1997) *Infect. Immun.* 65: 1109-1113); *Plasmodium falciparum* merozoite surface protein, MSP1 (Hui et al. (1996) *Infect. Immun.* 64: 1502- 1509); the malaria antigen Pf332 (Ahlborg et al. (1996) *Immunology* 88: 630-635); *Plasmodium falciparum* erythrocyte membrane protein I (Baruch et al. (1995) *Proc. Nat'l. Acad. Sci. USA* 93: 3497-3502; Baruch et al. (1995) *Cell* 82: 77-87); *Plasmodium falciparum* merozoite surface antigen, PfMSP-1 (Egan et al. (1996) *J Infect. Dis.* 173: 765- 769); *Plasmodium falciparum* antigens SERA, EBA- 175, RAP1 and RAP2 (Riley (1997) *J Pharm. Pharmacol.* 49: 21-27);
- 10 *Schistosoma japonicum* paramyosin (Sj97) or fragments thereof (Yang et al. (1995) *Biochem. Biophys. Res. Commun.* 212: 1029- 1039); and Hsp70 in parasites (Maresca and Kobayashi (1994) *Experientia* 50: 1067-1074).

15 2.9.6. CONTRACEPTION

Genetic vaccines that contain optimized antigens obtained by the methods of the invention are also useful for contraception. For example, genetic vaccines can be obtained that encode sperm cell specific antigens, and thus induce anti-

20 sperm immune responses. Vaccination can be achieved by, for example, administration of recombinant bacterial strains, e.g. *Salmonella* and the like, which express sperm antigen, as well as by induction of neutralizing anti-hCG antibodies by vaccination by DNA vaccines encoding human chorionic gonadotropin (hCG), or a fragment thereof.

25

Sperm antigens which can be used in the genetic vaccines include, for example, lactate dehydrogenase (LDH-C4), galactosyltransferase (GT), SP-10, rabbit sperm autoantigen (RSA), guinea pig (g)PH-20, cleavage signal protein (CS-1), HSA-63, human (h)PH-20, and AgX-1 (Zhu and Naz (1994) *Arch.*

Androl. 33: 141-144), the synthetic spenn peptide, P10G (O'Rand et al. (1993) J
Reprod. Immunol. 25: 89-102), the 135kD, 95kD, 65kD, 47kD, 41 kD and 23kD
proteins of sperm, and the FA-1 antigen (Naz et al. (1995) Arch. Androl. 35: 225-
23 1), and the 35 kD fragment of cytokeratin 1 (Lucas et al. (1996) Anticancer
5 Res. 16: 2493-2496).

The methods of the invention can also be used to obtain genetic vaccines
that are expressed specifically in testis. For example, polynucleotide sequences
that direct expression of genes that are specific to testis can be used (e.g.,
10 fertilization antigen-1 and the like). In addition to sperm antigens, antigens
expressed on oocytes or hormones regulating reproduction may be useful targets
of contraceptive vaccines. For example, genetic vaccines can be used to generate
antibodies against gonadotropin releasing hormone (GnRH) or zona pellucida
proteins (Miller et al. (1997) Vaccine 15:185 8-1862). Vaccinations using these
15 molecules have been shown to be efficacious in animal models (Miller et al.
(1997) Vaccine 15:1858-1862). Another example of a useful component of a
genetic contraceptive vaccine is the ovarian zona pellucida glycoprotein ZP3
(Tung et al. (1994) Reprod Fertil. Dev. 6:349-355).

20

2.10. MALARIAL ANTIGENS AND VACCINES

The present invention generally relates to the Plasmodium falciparum
erythrocyte membrane protein 1 ("PfEMP1"), nucleic acids which encode
25 PfEMP1, and antibodies which specifically recognize PfEMP1. The polypeptides,
antibodies and nucleic acids are useful in a variety of applications including
therapeutic, prophylactic, including vaccination, diagnostic and screening
applications.

The data described herein, indicates that PfEMP1 is responsible for both antigenic variation and receptor properties on PE, both of which are central to the special virulence and pathology of *P. falciparum*. The central role of PfEMP1 in *P. falciparum* biology, as the malarial adherence receptor for host proteins on microvascular endothelium, as described herein, indicates its usefulness in a malaria vaccine, in modelling prophylactic drugs, and also as a target for therapeutics to reverse PE adherence in acute cerebral malaria (Howard and Gilladoga, 1989).

2.10.1. MALARIAL POLYPEPTIDES

Soluble PfEMP1 has been reported to bind to CD36, TSP and ICAM-1, and tryptic fragments of PfEMP1 cleaved from the PE surface have been shown to bind to TSP or CD36 (Baruch, et al., Molecular Parasitology Meeting at Woods Hole, Sept 18- 22, 1994). Accordingly, in one aspect, the present invention provides substantially pure PfEMP1 polypeptides, analogs or biologically active fragments thereof.

The terms "substantially pure" or "isolated" refer, interchangeably, to proteins, polypeptides and nucleic acids which are separated from proteins or other contaminants with which they are naturally associated. A protein or polypeptide is considered substantially pure when that protein makes up greater than about 50% of the total protein content of the composition containing that protein, and typically, greater than about 60% of the total protein content. More typically, a substantially pure protein will make up from about 75 to about 90% of the total protein. Preferably, the protein will make up greater than about 90%, and more preferably, greater than about 95% of the total protein in the composition.

The term "biologically active fragment" as used herein, refers to portions of the proteins or polypeptides, e.g., a PfEMP1 derived polypeptide, which portions possess a particular biological activity, e.g., one or more activities found in a full length PfEMP1 polypeptide. For example, such biological activity may include the ability to bind a particular protein, substrate or ligand, to elicit antibodies reactive with PE, PfEMP1, the recombinant proteins or fragments thereof, to block, reverse or otherwise inhibit an interaction between two proteins, between an enzyme and its substrate, between an epitope and an antibody, or may include a particular catalytic activity. With regard to the polypeptides of the present invention, particularly preferred polypeptides or biologically active fragments include, e.g., polypeptides that possess one or more of the biological activities described above, such as the ability to bind a ligand of PfEMP1 or inhibit the binding of PfEMP1 to one or more of its ligands, e.g., CD36, TSP, ICAM-1, VCAM-1, ELAM-1, Chondroitin sulfate or by the presence within the polypeptide fragment of antigenic determinants which permit the raising of antibodies to that fragment.

The polypeptides of the present invention may also be characterized by their immunoreactivity with antibodies raised against PfEMP1 proteins or polypeptides. In particularly preferred aspects, the polypeptides are capable of inhibiting an interaction between a PfEMP1 protein and an antibody raised against a PfEMP1 protein. Additionally or alternatively, such fragments may be specifically immunoreactive with an antibody raised against a PfEMP1 protein. Such fragments are also referred to herein as "immunologically active fragments." Generally, such biologically active fragments will be from about 5 to about 500 amino acids in length.

Typically, these peptides will be from about 20 to about 250 amino acids in length, and preferably from about 50 to about 200 amino acids in length.

Generally, the length of the fragment may depend, in part, upon the application for which the particular peptide is to be used. For example, for raising antibodies, the peptides may be of a shorter length, e.g., from about 5 to about 50 amino acids in length, whereas for binding applications, the peptides may have a greater
5 length, e.g., from about 50 to about 500 amino acids in length, preferably, from about 100 to about 250 amino acids in length, and more preferably, from about 100 to about 200 amino acids in length.

The polypeptides of the present invention may generally be prepared using
10 recombinant or synthetic methods well known in the art. Recombinant techniques are generally described in Sambrook, et al., Molecular Cloning: A Laboratory Manual, (2nd ed.) Vols. 1-3, Cold Spring Harbor Laboratory, (1989). Techniques for the synthesis of polypeptides are generally described in Merrifield, J. Amer. Chem. Soc. 85:2149-2456 (1963), Atherton, et al., Solid Phase Peptide Synthesis:
15 A Practical Approach, IRL Press (1989), and. Merrifield, Science 232:341-347 (1986).

In preferred aspects, the polypeptides of the present invention may be expressed by a suitable host cell that has been transfected with a nucleic acid of
20 the invention, as described in greater detail below. Isolation and purification of the polypeptides of the present invention can be carried out by methods that are generally well known in the art. For example, the polypeptides may be purified using readily available chromatographic methods, e.g., ion exchange, hydrophobic interaction, HPLC or affinity chromatography, to achieve the desired
25 purity. Affinity chromatography may be particularly attractive in allowing the investigator to take advantage of the specific biological activity of the desired peptide, e.g., ligand binding, presence of antigenic determinants, or the like.

Exemplary polypeptides of the present invention will generally comprise an amino acid sequence that is substantially homologous to the amino acid sequence of a PfEMP1 protein, or biologically active fragments thereof, or may include sequences that may take on a homologous conformation. In particularly
5 preferred aspects, the polypeptides of the present invention will comprise an amino acid sequence that is substantially homologous to the amino acid sequence shown, described &/or referenced herein (including incorporated by reference), or a biologically active fragment thereof.

10 By "substantially homologous" is meant an amino acid sequence which is at least about 50% homologous to the amino acid sequence of PfEMP1 or a biologically active fragment thereof, preferably at least about 90% homologous, and more preferably at least about 95% homologous. In some aspects, substantially homologous may include a sequence that is at least 50%
15 homologous, but that presents a homologous structure in three dimensions, i.e., includes a substantially similar surface charge or presentation of hydrophobic groups.

Examples of preferred polypeptides include polypeptides having an amino
20 acid sequence substantially homologous to the MC PfEMP1 amino acid sequence as shown, described &/or referenced herein (including incorporated by reference), and PfEMP1 of other *P. falciparum* strains as shown, described &/or referenced herein (including incorporated by reference), as well as biologically active fragments of these polypeptides. Preferred peptides include those peptide
25 fragments of PfEMP1 that are involved in the sequestration of parasitized erythrocytes. Examples of these preferred peptides include peptides which comprise an amino acid sequence which is substantially homologous to amino acids 576 through 755 of the PfEMP1 amino acid sequence shown, described &/or referenced herein (including incorporated by reference).

Also among the particularly preferred peptides of the present invention are those peptides and peptide fragments of PfEMP1 which are relatively conserved among the variant strains of *P. falciparum* or which contain regions of high
5 homology to PfEMP1 proteins from other strains. The term "relatively conserved" generally refers to amino acid sequences that are substantially homologous to portions of the amino acid sequence shown, described &/or referenced herein (including incorporated by reference). However, also included within the definition of this term are peptides which are encoded by a nucleic acid which is a
10 PCR product of primer probes, and particularly, universal primers, derived from the PfEMP1 nucleic acid sequence. In particular, primer is probes derived from the nucleic acid sequence shown, described &/or referenced herein (including incorporated by reference), may be used to amplify nucleic acids from other strains of *P. falciparum*. Particularly preferred primer sequences include the
15 primer sequences shown in Table 1, below. Similarly, universal primer compositions, described in greater detail below and also shown in Table 1, may be used to amplify sequences that encode the peptides of the present invention.

Specific examples of relatively conserved peptides include those that are
20 contained in a region of PfEMP1 proteins that corresponds to amino acids 576 through 755 of the amino acid sequence of MC PfEMP1, as shown, described &/or referenced herein (including incorporated by reference).

Similar regions have been specifically elucidated in a number of *P. falciparum* strains (as described herein). In general, these corresponding regions
25 may be described as containing amino acid sequences that are encoded by the universal primer sequences described below. Generally, these amino acid sequences have one or more of the following general structures:

TTIDKX₁LX₂HE and/or FFWX₃WVX₄X₅ML

where X₁ is selected from leucine or isoleucine, X₂ is selected from glutamine and asparagine, X₃ is selected from the methionine, lysine and aspartic acid, X₄ is
5 selected from histidine, threonine and tyrosine and X₅ is selected from aspartic acid, glutamic acid and histidine. In particularly preferred aspects, the polypeptides may contain both of the above general amino acid sequences. Particularly preferred amino acid sequences will possess the conserved amino acids shown in the various fragments shown, described &/or referenced herein
10 (including incorporated by reference). In particular, conserved amino acid sequences of six amino acids or greater, shown, described &/or referenced herein (including incorporated by reference), may be used as epitopes for generation of antibodies that cross react with multiple *P. falciparum* strains.

15 The peptides of the invention may be free or tethered, or may include labeled groups for detection of the presence of the polypeptides. Suitable labels include radioactive, fluorescent and catalytic labeling groups that are well known in the art and that are substantially described herein, e.g., signaling enzymes, chemical reporter groups, polypeptide signals, biotin and the like. Additionally,
20 the peptides may include modifications to the N and C-termini of the peptide, e.g., an acylated N-terminus or amidated C-terminus.

Also included within the present invention are amino acid variants of the above described polypeptides. These variants may include insertions, deletions
25 and substitutions with other amino acids. For example, in some aspects, amino acids may be substituted with different amino acids having similar structural characteristics, e.g., net charge, hydrophobicity, or the like. For example, phenylalanine may be substituted with tyrosine, as a similarly hydrophobic

residue. Glycosylation modifications, either changed, increased amounts or decreased amounts, as well as other sequence modifications are also envisioned.

In addition to the above polypeptides which consist only of naturally-
5 occurring amino acids, peptidomimetics of the polypeptides of the present invention are also provided. Peptide analogs are commonly used in the pharmaceutical industry as non-peptide drugs with properties analogous to those of the template peptide. These types of non-peptide compound are termed "peptide mimetics" or "peptidomimetics" (Fauchere, J. (1986) Adv. Drug Res.
10 15:29; Veber and Freidinger (1985) TINS p.392; and Evans et al. (1987) J. Med. Chem 30:1229, and are usually developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to therapeutically useful peptides may be used to produce an equivalent therapeutic or prophylactic effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide
15 (i.e., a polypeptide that has a biological or pharmacological activity), such as naturally- occurring receptor-binding polypeptide, but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: -CH₂NH-, -CH₂S-, -CH₂-CH₂-, -CH=CH- (cis and trans), -COCH₂-, -CH(OH)CH₂-, and -CH₂SO-, by methods known in the art and further described
20 in the following references: Spatola, A.F. in Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983); Spatola, A.F., Vega Data (March 1983), Vol. 1, Issue 3, "Peptide Backbone Modifications" (general review); Morley, J.S., Trends Pharm Sci (1980) pp. 463-468 (general review); Hudson, D. et al., Int J Pept Prot Res
25 (1979) 14:177-185 (-CH₂NH-, CH₂CH₂-); Spatola, A.F. et al., Life Sci (1986) 38:1243-1249 (-CH₂-S); Hann, M.M., J Chem Soc Perkin Trans I (1982) 307-314 (-CH-CH-, cis and trans); Almquist, R.G. et al., J Med Chem (1980) 23:1392-1398 (-COCH₂-); Jennings-White, C. et al., Tetrahedron Lett (1982) 23:2533 (-COCH₂-); Szelke, M. et al., European Appln. EP 45665 (1982) CA: 97:39405

(1982) (-CH(OH)CH₂-); Holladay, M.W. et al., *Tetrahedron Lett* (1983) 24:4401-4404 (-C(OH)CH₂-); and Hruby, V.J., *Life Sci* (1982) 31:189-199 (-CH₂-S-)'

- Peptide mimetics may have significant advantages over polypeptide
embodiments, including, for example: more economical production, greater
5 chemical stability, enhanced pharmacological properties (half-life, absorption,
potency, efficacy, etc.), altered specificity (e.g., a broad-spectrum of biological
activities), reduced antigenicity, and others.

- Labeling of peptidomimetics usually involves covalent attachment of one
10 or more labels, directly or through a spacer (e.g., an amide group), to non-
interfering position(s) on the peptidomimetic that are predicted by quantitative
structure- activity data and/or molecular modeling. Such non-interfering positions
generally are positions that do not form direct contacts with the molecules to
which the peptidomimetic binds (e.g., CD36) to produce the therapeutic effect.
15 Derivatization (e.g., labeling) of peptidomimetics should not substantially
interfere with the desired biological or pharmacological activity of the
peptidomimetic. Generally, peptidomimetics of peptides of the invention bind to
their ligands (e.g., CD36) with high affinity and possess detectable biological
activity (i.e., are agonistic or antagonistic to one or more ligand-mediated
20 phenotypic changes).

- Systematic substitution of one or more amino acids of a consensus
sequence with a D-amino acid of the same type (e.g., D-lysine in place of L-
lysine) may be used to generate more stable peptides. In addition, constrained
25 peptides comprising a consensus sequence or a substantially identical consensus
sequence variation may be generated by methods known in the art (Rizo and
Gierasch (1992) *Ann. Rev. Biochem.* 61: 387; for example, by adding internal
cysteine residues capable of forming intramolecular disulfide bridges which
cyclize the peptide.

Polypeptides of the present invention may also be characterized by their ability to bind antibodies raised against PfEMP1, or fragments thereof. Preferably, these antibodies recognize polypeptide domains that are homologous to the

5 PfEMP1 proteins from a number of variants of *P. falciparum*. These homologous domains will generally be present throughout the family of PfEMP1 proteins. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein or domain. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies

10 specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Antibodies to PfEMP1 and its fragments are discussed in greater detail, below. As used herein, the terms "polypeptide" or

15 "peptide" are used interchangeably to refer to peptides, peptidomimetics, analogs, and the like, as described above.

The polypeptides of the present invention may be used as isolated polypeptides, or may exist as fusion proteins. A "fusion protein" generally refers

20 to a composite protein made up of two or more separate, heterologous proteins which are normally not fused together as a single protein.

Thus, a fusion protein may comprise a fusion of two or more heterologous or homologous sequences, provided these sequences are not normally fused

25 together. Fusion proteins will generally be made by either recombinant nucleic acid methods, i.e., as a result of transcription and translation of a gene fusion comprising a segment encoding a polypeptide comprising a PfEMP1 protein and a segment which encodes one or more heterologous proteins, or by chemical synthesis methods well known in the art.

2.10.2. MALARIAL NUCLEIC ACIDS AND CELLS CAPABLE OF EXPRESSING SAME

5

Also provided in the present invention are isolated nucleic acid sequences which encode the above described polypeptides and biologically active fragments. Typically, such nucleic acid sequences will comprise a segment that is substantially homologous to a portion or fragment of the nucleic acid sequence shown, described &/or referenced herein (including incorporated by reference). Preferably, the nucleic acids of the present invention will comprise at least about 15 consecutive nucleotides of the nucleic acid, more preferably, at least about 20 contiguous nucleotides, still more preferably, at least about 30 contiguous nucleotides, and still more preferably, at least about 50 contiguous nucleotides from the nucleotide sequence.

Substantial homology in the nucleic acid context means that the segments, or their complementary strands, when compared, are the same when properly aligned with the appropriate nucleotide insertions or deletions, in at least about 60% of the nucleotides, typically, at least about 70%, more typically, at least about 80%, usually, at least about 90%, and more usually, at least about 95% to 98% of the nucleotides. Alternatively, substantial homology exists when the segments will hybridize under selective hybridization conditions to a strand, or its complement, typically using a sequence of at least about 15 contiguous nucleotides derived from the PfEMP1 nucleic acid sequence. However, larger segments will usually be preferred, e.g., at least about 20 or contiguous nucleotides, more usually about 40 contiguous nucleotides, and preferably more than about 50 contiguous nucleotides. Selective hybridization exists when

hybridization occurs which is more selective than total lack of specificity. See, Kanchisa, Nucleic Acid Res. 12:203-213 (1984).

Nucleic acids of the present invention include RNA, cDNA, genomic
5 DNA, synthetic forms and mixed polymers, both sense and antisense strands. Furthermore, different alleles of each isoform are also included. The present invention also provides recombinant nucleic acids which are not otherwise naturally occurring. The nucleic acids included in the present invention will typically comprise RNA or DNA or mixed polymers. The DNA compositions will
10 generally include a coding region which encodes a polypeptide comprising an amino acid sequence substantially homologous to the amino acid sequence of a PfEMP1 protein. More preferred are those DNA segments comprising a nucleotide sequence which encodes a CD36 binding fragment of the PfEMP1 protein.

15

cDNA encoding the polypeptides of the present invention, or fragments thereof, may be readily employed as a probe useful for obtaining genes which encode the PfEMP1 polypeptides of the present invention. Preparation of these probes may be carried out by generally well known methods. For example, the
20 cDNA probes may be prepared from the amino acid sequence of the PfEMP1 protein. In particular, probes may be prepared based upon segments of the amino acid sequence which possess relatively low levels of degeneracy, i.e., few or one possible nucleic acid sequences which encode therefor.

25

Suitable synthetic DNA fragments may then be prepared, e.g., by the phosphoramidite method described by Beaucage and Carruthers, Tetra. Letts. 22:1859-1862 (1981). Alternatively, nucleotide sequences which are relatively conserved among the PfEMP1 coding sequences for the various *P. falciparum* strains may be used as suitable probes. A double stranded probe may then be

obtained by either synthesizing the complementary strand and hybridizing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence. Such cDNA probes may be used in the design of oligonucleotide probes and primers for
5 screening and cloning such genes, e.g., using well known PCR techniques, or, alternatively, may be used to detect the presence or absence of a PfEMP1 gene in a cell. Such nucleic acids, or fragments may comprise part or all of the cDNA sequence that encodes the polypeptides of the present invention. Effective cDNA probes may comprise as few as 15 consecutive nucleotides in the cDNA
10 sequence, but will often comprise longer segments. Further, these probes may further comprise an additional nucleotide sequence, such as a transcriptional primer sequence for cloning, or a detectable group for easy identification and location of complementary sequences.

15 cDNA or genomic libraries of various types may be screened for new alleles or related sequences using the above probes. The choice of cDNA libraries normally corresponds to tissue sources which are abundant in mRNA for the desired polypeptides. Phage libraries are normally preferred, e.g., λ gt11, but plasmid or YAC libraries may also be used. Clones of a library are spread onto
20 plates, transferred to a substrate for screening, denatured, and probed for the presence of the desired sequences.

In a related aspect, the nucleic acids of the present invention also include the PCR product or RT-PCR product, produced using the above described primer
25 probes. For example, primer probes derived from the nucleotide sequence shown, described &/or referenced herein (including incorporated by reference), may be used to amplify sequences from different malaria parasites, and in particular, different strains of *P. falciparum*.

The nucleic acids of the present invention may be present in whole cells, cell lysates or in partially pure or substantially pure or isolated form. Such "substantially pure" or "isolated" forms of these nucleic acids generally refer to the nucleic acid separated from contaminants with which it is generally associated, e.g., lipids, proteins and other nucleic acids. The nucleic acids of the present invention will be greater than about 50% pure. Typically, the nucleic acids will be more than about 60% pure, more typically, from about 75% to about 90% pure, and preferably, from about 95% to about 98% pure.

The present invention also provides substantially similar nucleic acid sequences, allelic variations and natural or induced sequences of the above described nucleic acids, as well as chemically modified and substituted nucleic acids, e.g., those which incorporate modified nucleotide bases or which incorporate a labeling group. In addition to comprising a segment which encodes a PfEMP1 protein or fragment thereof, the nucleic acids of the present invention may also comprise a segment encoding a heterologous protein, such that the gene is expressed to produce the two proteins as a fusion protein, as substantially described above.

In addition to their use as probes, the nucleic acids of the present invention may also be used in the preparation of the polypeptides of the present invention, as described above. DNA encoding the polypeptides of the present invention will typically be incorporated into DNA constructs capable of introduction to and expression in an *in vitro* cell culture. Often, the nucleic acids of the present invention may be used to produce a suitable recombinant host cell.

Specifically, DNA constructs will be suitable for replication in a unicellular host, such as bacteria, e.g., *E. coli*, viruses or yeast, but may also be intended for introduction into a cultured mammalian, plant, insect, or other

eukaryotic cell lines. DNA constructs prepared for introduction into bacteria or yeast will typically include a replication system recognized by the host, the intended DNA segment encoding the desired polypeptide, transcriptional and translational initiation and termination regulatory sequences operably linked to the polypeptide encoding segment. A DNA segment is operably linked when it is placed into a functional relationship with another DNA segment. For example, a promoter or enhancer is operably linked to a coding sequence if it stimulates the transcription of the sequence; DNA for a signal sequence is operably linked to DNA encoding a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide. Generally, DNA sequences that are operably linked are contiguous, and in the case of a signal sequence both contiguous and in reading phase. However, enhancers need not be contiguous with the coding sequences whose transcription they control. Linking is accomplished by ligation at convenient restriction sites or at adapters or linkers inserted in lieu thereof. The selection of an appropriate promoter sequence will generally depend upon the host cell selected for the expression of the DNA segment.

Examples of suitable promoter sequences include prokaryotic, and eukaryotic promoters well known in the art. See, e.g., Sambrook et al., *supra*. The transcriptional regulatory sequences will typically include a heterologous enhancer or promoter which is recognized by the host. The selection of an appropriate promoter will depend upon the host, but promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters are known and available. See Sambrook et al., *supra*.

Conveniently available expression vectors which include the replication system and transcriptional and translational regulatory sequences together with the insertion site for the PfEMP1 polypeptide encoding segment may be employed. Examples of workable combinations of cell lines and expression

vectors are described in Sambrook et al., supra, and in Metzger et al., Nature 334:31-36 (1988).

5 The vectors containing the DNA segments of interest, e.g., those encoding polypeptides comprising a PfEMP1 protein or fragments thereof, can be transferred into the host cell by well known methods, which may vary depending upon the type of host used. For example, calcium chloride transfection is commonly used for prokaryotic cells, whereas calcium phosphate treatment may be used for other hosts. See, Sambrook et al., supra. The term "transformed cell"
10 as used herein, includes the progeny of originally transformed cells.

Techniques for manipulation of nucleic acids which encode the polypeptides of the present invention, i.e., subcloning the nucleic acids into expression vectors, labeling probes, DNA hybridization and the like, are generally
15 described in Sambrook, et al., supra. In recombinant methods, generally the nucleic acid encoding a peptide of the present invention is first cloned or isolated in a form suitable for ligation into an expression vector. After ligation, the vectors containing the nucleic acids fragments or inserts are introduced into a suitable host cell, for the expression of the polypeptide of the invention. The polypeptides
20 may then be purified or isolated from the host cells. Methods for the synthetic preparation of oligonucleotides are generally described in Gait, oligonucleotide Synthesis: A Practical Approach, IRL Press (1990).

There are various methods of isolating the nucleic acids which encode the
25 polypeptides of the present invention. Typically, the DNA is isolated from a genomic or cDNA library using labeled oligonucleotide probes specific for sequences in the desired DNA. Restriction endonuclease digestion of genomic DNA or cDNA containing the appropriate genes can be used to isolate the DNA encoding the binding domains of these proteins. From the PfEMP1 sequence

given (as shown herein), a panel of restriction endonucleases can be constructed to give cleavage of the DNA in desired regions, i.e., to obtain segments which encode biologically active fragments of the PfEMP1 protein. Following restriction endonuclease digestion, DNA encoding the polypeptides of the present invention
5 is identified by its ability to hybridize with a nucleic acid probe in, for example a Southern blot format. These regions are then isolated using standard methods. See, e.g., Sambrook, et al., supra.

The polymerase chain reaction, or "PCR" can also be used to prepare
10 nucleic acids which encode the polypeptides of the present invention. PCR technology is used to amplify nucleic acid sequences of the desired nucleic acid, e.g., the DNA which encodes the polypeptides of the invention, directly from mRNA, cDNA, or genomic or cDNA libraries.

15 Appropriate primers and probes for amplifying the nucleic acids described herein, may be generated from analysis of the PfEMP1 oligonucleotide sequence, such as those shown, described &/or referenced herein (including incorporated by reference) and Table 1. Briefly, oligonucleotide primers complementary to the two 31 borders of the DNA region to be amplified are synthesized. The PCR is then
20 carried out using the two primers. See, e.g., PCR Protocols: A Guide to Methods and Applications (Innis, M., Gelfand, D., Sninsky, J. and White, T., eds.) Academic Press (1990). Primers can be selected to amplify various sized segments from the PfEMP1 oligonucleotide sequence. The primers may also contain a restriction site and additional bases to permit "in-frame" cloning of the
25 insert into an appropriate expression vector, using the restriction sites present on the primers.

2.10.3. ANTIBODIES

The nucleic acids and polypeptides of the present invention, or fragments thereof, are also useful in producing antibodies, either polyclonal or monoclonal. These antibodies are produced by immunizing an appropriate vertebrate host, e.g., rat, mouse, rabbit or goat, with a polypeptide of the invention, or its fragment, or plasmid DNA containing a nucleic acid of the invention, alone or in conjunction with an adjunct. Usually, two or more immunizations are involved, and a few days following the last injection, the blood or spleen of the host will be harvested.

For production of polyclonal antibodies, an appropriate target immune system is selected, typically a mouse or rabbit, but also including goats, sheep, cows, guinea pigs, monkeys and rats. The substantially purified antigen or plasmid is presented to the immune system in a fashion determined by methods appropriate for the animal. These and other parameters are well known to immunologists. Typically, injections are given in the footpads, intramuscularly, intradermally or intraperitoneally. The immunoglobulins produced by the host can be precipitated, isolated and purified by routine methods, including affinity purification.

For monoclonal antibodies, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of these animals are excised and individual spleen cells are fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone are tested for the production of an appropriate antibody specific for the desired region of the antigen. Techniques for producing antibodies are well known in the art. See, e.g., Goding et al., *Monoclonal Antibodies: Principles and Practice* (2d ed.)

- Acad. Press, N.Y., and Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York (1988). Other suitable techniques involve the *in vitro* exposure of lymphocytes to the antigenic polypeptides or alternatively, to selection of libraries of antibodies in phage or similar vectors. Huse et al.,
- 5 Generation of Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda, Science 246:1275-1281 (1989). Monoclonal antibodies with affinities of 10^8 liters/mole, preferably 10^9 to 10^{10} or stronger, will be produced by these methods.
- 10 The antibodies generated can be used for a number of purposes, e.g., as probes in immunoassays, for inhibiting PfEMP1 binding to its ligands, thereby inhibiting or reducing erythrocyte sequestration, in diagnostics or therapeutics, or in research to further elucidate the mechanism of various aspects of malarial infection, and particularly, *P. falciparum* infection. The antibodies of the present
- 15 invention can be used with or without modification. Frequently, the antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. Such labels include those that are well known in the art, such as the labels described previously for the polypeptides of the invention. Additionally, the antibodies of the invention may be chimeric, human-
- 20 like or humanized, in order to reduce their potential antigenicity, without reducing their affinity for their target. Chimeric, human-like and humanized antibodies have generally been described in the art. Generally, such chimeric, human-like or humanized antibodies comprise variable regions, e.g., complementarity determining regions (CDR) (for humanized antibodies), from a mammalian
- 25 animal, i.e., a mouse, and a human framework region. By incorporating as little foreign sequence as possible in the hybrid antibody, the antigenicity is reduced. Preparation of these hybrid antibodies may be carried out by methods well known in the art.

Preferred antibodies are those that are specifically immunoreactive with the polypeptides of the present invention and their immunologically active fragments. The phrase "specifically immunoreactive," when referring to the interaction between an antibody of the invention and a particular protein, refers to an antibody that specifically recognizes and binds with relatively high affinity to the particular protein, such that this binding is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

The antibodies generated can be used for a number of purposes, e.g., as probes in immunoassays, for inhibiting interaction between a PfEMP1 protein and its ligand, e.g., CD-36, TSP, ICAM-1, VCAM-1, ELAM-1, or Chondroitin sulfate, thereby inhibiting or reducing the level of PfEMP1-ligand interaction, in diagnostics or therapeutics, or in research to further elucidate the mechanism of malarial pathology, e.g., erythrocyte sequestration. Where the antibodies are used to block or reverse the interaction between a polypeptide of the invention and an associating ligand or PE, the antibody will generally be referred to as a "blocking antibody." Preferred antibodies are those monoclonal or polyclonal antibodies which specifically recognize and bind the polypeptides of the invention. Accordingly, these preferred antibodies will specifically recognize and bind the polypeptides which have an amino acid sequence that is substantially homologous

to the relevant amino acid sequence shown, described &/or referenced herein (including incorporated by reference), or immunologically active fragments thereof. Still more preferred are antibodies which are capable of forming an antibody-ligand complex with the relatively conserved polypeptide fragments of PfEMP1 sequences, and are thereby capable of blocking an interaction of PfEMP1 from a variety of *P. falciparum* strains, and PfEMP1 ligands.

2.10.4. METHODS OF USE

The polypeptides, antibodies, and nucleic acids of the present invention have a variety of important uses, including, but not limited to, diagnostic, screening, prophylactic, including vaccination, and therapeutic applications.

2.10.4.1. DIAGNOSTIC APPLICATIONS

In a particularly preferred aspect, the present invention provides methods and reagents useful in detecting the presence of PfEMP1 in a sample. These detection methods are particularly useful in diagnosing malarial infections in a patient. For example, in a particularly preferred aspect, the antibodies of the present invention may be used to assay for the presence or absence of PfEMP1 in a sample. Immunoassay techniques for the detection of the particular antigen are very well known in the art. For a review of immunological and immunoassay procedures in general, see Basic and Clinical Immunology 7th Edition (D. Stites and A. Terr ed.) 1991.

Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Enzyme Immunoassay, E.T. Maggio, ed., CRC Press, Boca Raton, Florida (1980);

"Practice and Theory of Enzyme Immunoassays," P. Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers B.V. Amsterdam (1985); and, Harlow and Lane, Antibodies, A Laboratory Manual, supra. Generally, these methods comprise contacting the antibody with a sample to be tested, and detecting any specific binding between the antibody and a protein within the sample. Typically, this will be in a blot format, e.g., western blot, or in an ELISA format. Methods of performing these assay formats are well known in the art. See, e.g., Basic and Clinical Immunology, 7th ed. (D. Stites and A Terr, eds., 1991).

10

Typically, these diagnostic methods comprise contacting a sample with an antibody to PfEMP1, as described herein, and determining whether the antibody binds to any portion of the sample. In the case of human diagnostic techniques, the sample may be a whole blood sample, or some fraction thereof, e.g. an erythrocyte containing sample. Generally, such diagnostic methods are well known in the art, and are described in the above described references. The immunoreactivity of the antibody with the sample, indicates the presence of PfEMP1 in the sample, and, in the case of a sample derived from a patient, a possible malarial infection.

20

Alternatively, labeled polypeptides of the present invention may be used as diagnostic reagents in detecting the presence or absence of antibodies to PfEMP1, in a patient. The presence of antibodies within a patient would be indicative that the patient had been exposed to a malaria parasite sufficiently to result in an antigenic response.

25

Similarly, the nucleic acid probes of the invention may be used in a similar manner, i.e., to identify the presence in a sample of a DNA segment encoding a PfEMP1 polypeptide, or as PCR or RT-PCR primers to amplify and then detect

PfEMP1 encoding nucleic acid segments. Such assays typically involve the immobilization of nucleic acids in the sample, followed by interrogation?? of the immobilized sequences with a chemically labeled oligonucleotide probe, as described herein. Hybridization of the probe to the immobilized sample indicates
5 the presence of a DNA segment encoding PfEMP1, and thus, a malarial infection. As described above, assays may be further designed to indicate not only the presence of a Malarial parasite, but also indicate the strain of parasite present. Although described in terms of an immobilized sample probed with a solution based oligonucleotide probe, a wide variety of assay conformations may be
10 adopted, which conformations are generally well known in the art.

15

2.10.4.2. SCREENING APPLICATIONS

In another particularly preferred aspect, the present invention provides methods for screening compounds to determine whether or not the particular
20 compound is an antagonist of a symptom of a malarial infection. In particular, the screening methods of the present invention can be used to determine whether a test compound is an antagonist of the sequestration of erythrocytes which is associated with *P. falciparum* malaria. More particularly, the screening methods can determine whether a compound is an antagonist of the PfEMP1/ligand
25 interaction. Ligands of PfEMP1 generally include, e.g., CD36, TSP, ELAM-1, ICAM-1, VCAM-1 or Chondroitin sulfate.

Generally, the screening methods of the present invention comprise contacting PfEMP1 protein, or a fragment thereof, and/or ligand protein, with a

compound which is to be screened ("test compound"). The level of PfEMP1/ligand complex formed may then be detected and compared to a control, e.g., in the absence of the test compound. A decrease in the level of PfEMP1/ligand interaction is indicative that the test compound is an antagonist of
5 that interaction.

A test compound may be a chemical compound, a mixture of chemical compounds, a biological macromolecule, or an extract made from biological materials, such as bacteria, phage, yeast, plants, fungi, animal cells or tissues. Test
10 compounds are evaluated for potential activity as antagonists of PfEMP1/ligand interaction by inclusion in the screening assays described herein. An "antagonist" refers to a compound which will diminish the level of PfEMP1/ligand interaction, over a control.

15 It will often be desirable in the screening assays of the present invention, to provide one of the PfEMP1 or ligand proteins immobilized on a solid support. Suitable solid supports include, e.g., agarose, cellulose, dextran, Sephadex, Sepharose, carboxymethyl cellulose, polystyrene, filter paper, nitrocellulose, ion exchange resins, plastic films, glass beads, polyaminemethylvinylether maleic
20 acid copolymer, amino acid copolymer, ethylene-maleic acid copolymer, nylon, silk, etc. The support may be in the form of, e.g., a test tube, microtiter plate, beads, test strips, flat surface, e.g., for blotting formats, or the like. The reaction of the PfEMP1 polypeptide or its ligand with the particular solid support may be carried out by methods well known in the art, e.g., binding to an immobilized
25 anti-PfEMP1 antibody, or binding to prederivatized solid support.

In addition to the foregoing, it may also be desirable to provide either the PfEMP1 or its ligand linked to a suitable detectable group to make detection of binding of one protein to the other, simpler. Useful detectable groups, or labels,

are generally well known in the art. For example, a detectable group may be a radiolabel, such as, ^{125}I , ^{32}P or ^{35}S , or a fluorescent or chemiluminescent group.

Alternatively, the detectable group may be a substrate, cofactor, inhibitor, affinity ligand, antibody binding epitope tag, or an enzyme which is capable of being assayed. Suitable enzymes include, e.g., horseradish peroxidase, luciferase, or another readily assayable enzymes. These enzyme groups may be attached to the PfEMP1 polypeptide, or its ligand by chemical means or maybe expressed as a fusion protein, as already described.

10

Generally, where one of the above proteins, e.g., the PfEMP1 ligand, is immobilized on a solid support, the other protein, e.g., PfEMP1 or its fragment, will be labeled with an appropriate detectable group. Assaying whether a compound is an antagonist of the interaction of the two proteins is then a matter of contacting the labeled PfEMP1 polypeptide or fragment with the immobilized ligand, in the presence of the test compound, under conditions which allow specific binding of the two proteins. The amount of label bound to the solid support is compared to a control, where no test compound was added. Where a test compound results in a reduction of the amount of label which binds to a solid support, that compound is an antagonist of the PfEMP1/ligand interaction.

15
20

25 2.10.4.3. THERAPEUTIC AND PROPHYLACTIC APPLICATIONS

In addition to the above described uses, the polypeptides of the present invention may also be used in therapeutic applications, for the treatment of human and/or non-human mammalian patients. The therapeutic uses of the polypeptides

of the present invention include the treatment of symptoms of existing disorders, as well as prophylactic applications. The term "prophylactic" refers to the prevention of a particular disorder, or symptoms of a particular disorder. Thus, prophylactic treatments will generally include drugs which actively participate in the prevention of a particular disorder such as a malaria infection, or symptoms thereof. Prophylactic applications will also include treatments which elicit a preventative response from a patient, including, for example, an immunological response as in the case of vaccination.

Typically, both therapeutic and prophylactic applications will comprise administering an effective amount of the compositions of the present invention to a patient, to treat or prevent symptoms, or the onset of a malarial parasite infection. An "effective amount", as the term is used herein, is defined as the amount of the composition which is necessary to achieve the desired goal, i.e. alleviation of symptoms, prevention of symptoms or infection, or treatment of disease.

In prophylactic applications, the polypeptides of the present invention may be used in a variety of treatments. For example, the polypeptides of the invention are particularly useful as a vaccine, to elicit an immunological response by a patient, e.g., production of antibodies specific for PfEMP1. In particular, such vaccine applications generally involve the administration of the PfEMP1 protein or biologically active fragments thereof, to the host or patient.

In response to this administration, the patient's immune system will generate antibodies to the particular PfEMP1 protein or fragment introduced. An amount of the polypeptides sufficient to produce an immunological response in a patient is termed "an immunogenically effective amount." Thus, the vaccines of the present invention will contain an immunogenically effective amount of the

polypeptides of the present invention. The immune response of the patient may include generation of antibodies, activation of cytotoxic T- lymphocytes against cells expressing the polypeptides, e.g., PE, or other mechanisms known to the skilled artisan. See, e.g., Paul, Fundamental Immunology, 2d Edition, Raven Press. Useful carriers are well known in the art, and include for example, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, polyamino acids such as poly(D-lysine; D- glutamic acid), influenza, hepatitis B virus core protein, hepatitis B virus recombinant vaccine. The vaccines can also contain a physiologically tolerable diluent, such as water, buffered water, buffered saline, saline and typically may further include an adjuvant, such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, alum, or other materials well known in the art.

Alternatively, the nucleic acids of the present invention may also be used as vaccines for the prevention of malaria symptoms, and/or infection by malaria parasites. See Sedegah, et al. Proc. Nat'l Acad. Sci. (1994) 91:9866-9870.

For example, plasmid DNA comprising the nucleic acids of the present invention may be directly administered to a patient. Expression of this "naked" DNA will have effects similar to the injection of the actual polypeptides, as described above. Specifically, the patient's immune response to the presence of the proteins expressed from the DNA, will result in the production of antibodies to that protein. The nucleic acids may also be used to design antisense probes to interrupt transcription of PfEMP1 peptides in parasitized erythrocytes.

Antisense methods are generally well known in the art. The polypeptides of the present invention, and analogs thereof, may also be used as prophylactic treatments to prevent the onset of symptoms of malarial infection. For example, administration of the polypeptides can directly inhibit, block or reverse the

sequestration of erythrocytes in patients suffering from *P. falciparum* malaria infections. In particular, the polypeptides of the invention may be used to compete with or displace PE associated PfEMP1 in binding CD36.

- 5 The blockage or reversal of sequestration will reduce or eliminate the microvascular occlusion generally associated with the pathology of this type of malaria, which, again, can lead to destruction of the PE by the host. The antibodies of the invention may also be used in a similar fashion. In particular, the antibodies, which are capable of binding the polypeptides of the present
- 10 invention, may be directly administered to a patient. By binding PfEMP1, the antibodies of the present invention are effective in blocking, reducing or reversing PfEMP1 mediated interactions, e.g., erythrocyte sequestration. Chimeric, human-like or humanized antibodies are particularly useful for administration to human patients. Additionally, such antibodies may also be used as a passive vaccination
- 15 method to provide a subject with a short term immunization, much as anti-hepatitis A injections have been used previously.

- In alternative aspects, the polypeptides, antibodies and nucleic acids of the invention may be used to treat a patient already suffering from a malarial
- 20 infection. In particular, the compositions of the present invention may be administered to a patient suffering from a malarial infection to treat symptoms associated with that infection. More particularly, these compositions may be administered to the patient to prevent or reduce erythrocyte sequestration and the resulting microvascular occlusion associated with malarial, and more specifically,
- 25 *P. falciparum*, infections.

Although the polypeptides, nucleic acids and antibodies of the present invention may be administered alone, for therapeutic and prophylactic applications, these elements will generally be administered as part of a

pharmaceutical composition, e.g., in combination with a pharmaceutically acceptable carrier. Typically, a single composition may be used in both therapeutic and prophylactic applications. Pharmaceutical formulations suitable for use in the present invention are generally described in Remington's Pharmaceutical
5 Sciences, Mack Publishing Co., 17th ed. (1985).

The pharmaceutical compositions of the present invention are intended for parenteral, topical, oral, or local administration. Where the pharmaceutical compositions are administered parenterally, the invention provides pharmaceutical
10 compositions that comprise a solution of the agents described above, e.g., polypeptides of the invention, dissolved or suspended in a pharmaceutically acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, saline glycine, and the like. These compositions may be sterilized by conventional, well known methods, e.g., sterile
15 filtration. The resulting aqueous solutions may be packaged for use as is, or lyophilized for combination with a sterile solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents, and the like, for
20 example sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose starch,
25 magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition may be formed by incorporating any of the normally employed excipients, such as the previously listed carriers, and generally, 10-95% of active ingredient, and more preferably 25-75% active

ingredient. In addition, for oral administration of peptide based compounds, the pharmaceutical compositions may include the active ingredient as part of a matrix to prevent proteolytic degradation of the active ingredient by digestive process, e.g., by providing the pharmaceutical composition within a liposomal
5 composition, according to methods well known in the art. See, e.g., Remington's Pharmaceutical Sciences, Mack Publishing Co., 17th Ed. (1985).

For aerosol administration, the polypeptides are generally supplied in finely divided form along with a surfactant or propellant. Preferably, the
10 surfactant will be soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids, with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. A carrier can also be
15 included, as desired, as with, e.g., lecithin for intranasal delivery. The above described compositions are suitable for a single administration or a series of administrations. When given as a series, e.g., as a vaccine booster, the inoculations subsequent to the initial administration are given to boost the immune response, and are typically referred to as booster inoculations.

20

The amount of the above compositions to be administered to the patient will vary depending upon what is to be administered to the patient, the state of the patient, the manner of administration, and the particular application, e.g., therapeutic or prophylactic. In therapeutic applications, the compositions are
25 administered to the patient already suffering from a malarial infection, in an amount sufficient to inhibit the spread of the parasite through the erythrocytes, and thereby cure or at least partially arrest the symptoms of the disease and its associated complications.

An amount adequate to accomplish this is termed "a therapeutically effective amount." Amounts effective for this use will depend upon the severity of the disease and the weight and general state of the patient, but will generally be in the range of from about 1 mg to about 5 g of active agent per day, preferably from
5 about 50 mg per day to about 500 mg per day, and more preferably, from about 50 mg to about 100 mg per day, for a 70 kg patient.

For prophylactic applications, immunogenically effective amounts will also depend upon the composition, the manner of administration and the weight
10 and general state of the patient, as well as the judgment of the prescribing physician. For the peptide, peptide analog and antibody based pharmaceutical compositions, the general range for the initial immunization (for either prophylactic or therapeutic applications) will be from about 100 μ g to about 1 g of polypeptide for a 70 kg patient, followed by boosting dosages of from about 1 μ g
15 to about 1 gm of polypeptide pursuant to a boosting regimen over weeks to months, depending upon the patient's response and condition, e.g., by measuring the level of parasite or antibodies in the patient's blood. For nucleic acids, typically from about 30 to about 100 μ g of nucleic acid is injected into a 70 kg patient, more typically, about 50 to 150 μ g of nucleic acid is injected, followed by
20 boosting treatments as appropriate.

The present invention is further illustrated by the following examples. These examples are merely to illustrate aspects of the present invention and are not intended as limitations of this invention.

25

2.11. DIRECTED EVOLUTION METHODS

In one aspect the invention described herein is directed to the use of
5 repeated cycles of reductive reassortment, recombination and selection which
allow for the directed molecular evolution of highly complex linear sequences,
such as DNA, RNA or proteins thorough recombination.

In vivo shuffling of molecules can be performed utilizing the natural
10 property of cells to recombine multimers. While recombination *in vivo* has
provided the major natural route to molecular diversity, genetic recombination
remains a relatively complex process that involves 1) the recognition of
homologies; 2) strand cleavage, strand invasion, and metabolic steps leading to
the production of recombinant chiasma; and finally 3) the resolution of chiasma
15 into discrete recombined molecules. The formation of the chiasma requires the
recognition of homologous sequences.

In a preferred embodiment, the invention relates to a method for producing
a hybrid polynucleotide from at least a first polynucleotide and a second
20 polynucleotide. The present invention can be used to produce a hybrid
polynucleotide by introducing at least a first polynucleotide and a second
polynucleotide which share at least one region of partial sequence homology into
a suitable host cell. The regions of partial sequence homology promote processes
which result in sequence reorganization producing a hybrid polynucleotide. The
25 term "hybrid polynucleotide", as used herein, is any nucleotide sequence which
results from the method of the present invention and contains sequence from at
least two original polynucleotide sequences. Such hybrid polynucleotides can
result from intermolecular recombination events which promote sequence
integration between DNA molecules. In addition, such hybrid polynucleotides

can result from intramolecular reductive reassortment processes which utilize repeated sequences to alter a nucleotide sequence within a DNA molecule.

The invention provides a means for generating hybrid polynucleotides which may encode biologically active hybrid polypeptides. In one aspect, the original polynucleotides encode biologically active polypeptides. The method of the invention produces new hybrid polypeptides by utilizing cellular processes which integrate the sequence of the original polynucleotides such that the resulting hybrid polynucleotide encodes a polypeptide demonstrating activities derived from the original biologically active polypeptides. For example, the original polynucleotides may encode a particular enzyme from different microorganisms. An enzyme encoded by a first polynucleotide from one organism may, for example, function effectively under a particular environmental condition, e.g. high salinity. An enzyme encoded by a second polynucleotide from a different organism may function effectively under a different environmental condition, such as extremely high temperatures. A hybrid polynucleotide containing sequences from the first and second original polynucleotides may encode an enzyme which exhibits characteristics of both enzymes encoded by the original polynucleotides. Thus, the enzyme encoded by the hybrid polynucleotide may function effectively under environmental conditions shared by each of the enzymes encoded by the first and second polynucleotides, e.g., high salinity and extreme temperatures.

Enzymes encoded by the original polynucleotides of the invention include, but are not limited to; oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases. A hybrid polypeptide resulting from the method of the invention may exhibit specialized enzyme activity not displayed in the original enzymes. For example, following recombination and/or reductive reassortment of polynucleotides encoding hydrolase activities, the resulting hybrid polypeptide

encoded by a hybrid polynucleotide can be screened for specialized hydrolase activities obtained from each of the original enzymes, i.e. the type of bond on which the hydrolase acts and the temperature at which the hydrolase functions. Thus, for example, the hydrolase may be screened to ascertain those chemical
5 functionalities which distinguish the hybrid hydrolase from the original hydrolyases, such as: (a) amide (peptide bonds), i.e. proteases; (b) ester bonds, i.e. esterases and lipases; (c) acetals, i.e., glycosidases and, for example, the temperature, pH or salt concentration at which the hybrid polypeptide functions.

10 Sources of the original polynucleotides may be isolated from individual organisms ("isolates"), collections of organisms that have been grown in defined media ("enrichment cultures"), or, most preferably, uncultivated organisms ("environmental samples"). The use of a culture-independent approach to derive polynucleotides encoding novel bioactivities from environmental samples is most
15 preferable since it allows one to access untapped resources of biodiversity.

"Environmental libraries" are generated from environmental samples and represent the collective genomes of naturally occurring organisms archived in cloning vectors that can be propagated in suitable prokaryotic hosts. Because the
20 cloned DNA is initially extracted directly from environmental samples, the libraries are not limited to the small fraction of prokaryotes that can be grown in pure culture. Additionally, a normalization of the environmental DNA present in these samples could allow more equal representation of the DNA from all of the species present in the original sample. This can dramatically increase the
25 efficiency of finding interesting genes from minor constituents of the sample which may be under-represented by several orders of magnitude compared to the dominant species.

For example, gene libraries generated from one or more uncultivated microorganisms are screened for an activity of interest. Potential pathways encoding bioactive molecules of interest are first captured in prokaryotic cells in the form of gene expression libraries. Polynucleotides encoding activities of interest are isolated from such libraries and introduced into a host cell. The host cell is grown under conditions which promote recombination and/or reductive reassortment creating potentially active biomolecules with novel or enhanced activities.

The microorganisms from which the polynucleotide may be prepared include prokaryotic microorganisms, such as Eubacteria and Archaeobacteria, and lower eukaryotic microorganisms such as fungi, some algae and protozoa. Polynucleotides may be isolated from environmental samples in which case the nucleic acid may be recovered without culturing of an organism or recovered from one or more cultured organisms. In one aspect, such microorganisms may be extremophiles, such as hyperthermophiles, psychrophiles, psychrotrophs, halophiles, barophiles and acidophiles. Polynucleotides encoding enzymes isolated from extremophilic microorganisms are particularly preferred. Such enzymes may function at temperatures above 100°C in terrestrial hot springs and deep sea thermal vents, at temperatures below 0°C in arctic waters, in the saturated salt environment of the Dead Sea, at pH values around 0 in coal deposits and geothermal sulfur-rich springs, or at pH values greater than 11 in sewage sludge. For example, several esterases and lipases cloned and expressed from extremophilic organisms show high activity throughout a wide range of temperatures and pHs.

Polynucleotides selected and isolated as hereinabove described are introduced into a suitable host cell. A suitable host cell is any cell which is capable of promoting recombination and/or reductive reassortment. The selected

polynucleotides are preferably already in a vector which includes appropriate control sequences. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or preferably, the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the
5 construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis et al, 1986).

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Salmonella typhimurium*; fungal
10 cells, such as yeast; insect cells such as *Drosophila S2* and *Spodoptera Sf9*; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; and plant cells. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

15 With particular references to various mammalian cell culture systems that can be employed to express recombinant protein, examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in "SV40-transformed simian cells support the replication of early SV40 mutants" (Gluzman, 1981), and other cell lines capable of expressing a
20 compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived
25 from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Host cells containing the polynucleotides of interest can be cultured in conventional nutrient media modified as appropriate for activating promoters,

selecting transformants or amplifying genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan. The clones which are identified as having the specified enzyme activity may then be
5 sequenced to identify the polynucleotide sequence encoding an enzyme having the enhanced activity.

In another aspect, it is envisioned the method of the present invention can be used to generate novel polynucleotides encoding biochemical pathways from
10 one or more operons or gene clusters or portions thereof. For example, bacteria and many eukaryotes have a coordinated mechanism for regulating genes whose products are involved in related processes. The genes are clustered, in structures referred to as "gene clusters," on a single chromosome and are transcribed together under the control of a single regulatory sequence, including a single
15 promoter which initiates transcription of the entire cluster. Thus, a gene cluster is a group of adjacent genes that are either identical or related, usually as to their function. An example of a biochemical pathway encoded by gene clusters are polyketides. Polyketides are molecules which are an extremely rich source of bioactivities, including antibiotics (such as tetracyclines and erythromycin), anti-
20 cancer agents (daunomycin), immunosuppressants (FK506 and rapamycin), and veterinary products (monensin). Many polyketides (produced by polyketide synthases) are valuable as therapeutic agents. Polyketide synthases are multifunctional enzymes that catalyze the biosynthesis of an enormous variety of carbon chains differing in length and patterns of functionality and cyclization.
25 Polyketide synthase genes fall into gene clusters and at least one type (designated type I) of polyketide synthases have large size genes and enzymes, complicating genetic manipulation and *in vitro* studies of these genes/proteins.

The ability to select and combine desired components from a library of polyketides, or fragments thereof, and postpolyketide biosynthesis genes for generation of novel polyketides for study is appealing. The method of the present invention makes it possible to facilitate the production of novel polyketide
5 synthases through intermolecular recombination.

Preferably, gene cluster DNA can be isolated from different organisms and ligated into vectors, particularly vectors containing expression regulatory sequences which can control and regulate the production of a detectable protein or
10 protein-related array activity from the ligated gene clusters. Use of vectors which have an exceptionally large capacity for exogenous DNA introduction are particularly appropriate for use with such gene clusters and are described by way of example herein to include the f-factor (or fertility factor) of *E. coli*. This f-factor of *E. coli* is a plasmid which affect high-frequency transfer of itself during
15 conjugation and is ideal to achieve and stably propagate large DNA fragments, such as gene clusters from mixed microbial samples. Once ligated into an appropriate vector, two or more vectors containing different polyketide synthase gene clusters can be introduced into a suitable host cell. Regions of partial sequence homology shared by the gene clusters will promote processes which
20 result in sequence reorganization resulting in a hybrid gene cluster. The novel hybrid gene cluster can then be screened for enhanced activities not found in the original gene clusters.

Therefore, in a preferred embodiment, the present invention relates to a
25 method for producing a biologically active hybrid polypeptide and screening such a polypeptide for enhanced activity by:

- 1) introducing at least a first polynucleotide in operable linkage and a second polynucleotide in operable linkage, said at least first polynucleotide and second polynucleotide sharing at least one region of partial sequence homology, into a suitable host cell;
- 2) growing the host cell under conditions which promote sequence reorganization resulting in a hybrid polynucleotide in operable linkage;
- 3) expressing a hybrid polypeptide encoded by the hybrid polynucleotide;
- 4) screening the hybrid polypeptide under conditions which promote identification of enhanced biological activity; and
- 5) isolating the a polynucleotide encoding the hybrid polypeptide.

Methods for screening for various enzyme activities are known to those of skill in the art and discussed throughout the present specification. Such methods may be employed when isolating the polypeptides and polynucleotides of the present invention.

As representative examples of expression vectors which may be used there may be mentioned viral particles, baculovirus, phage, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral DNA (e.g. vaccinia, adenovirus, fowl pox virus, pseudorabies and derivatives of SV40), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for specific hosts of interest (such as bacillus, aspergillus and yeast). Thus, for example, the DNA may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences. Large numbers of suitable vectors are known to those of skill in the art, and are commercially

- available. The following vectors are provided by way of example; Bacterial: pQE vectors (Qiagen), pBluescript plasmids, pNH vectors, (lambda-ZAP vectors (Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia).
- 5 However, any other plasmid or other vector may be used as long as they are replicable and viable in the host. Low copy number or high copy number vectors may be employed with the present invention.

- A preferred type of vector for use in the present invention contains an
- 10 f-factor origin replication. The f-factor (or fertility factor) in *E. coli* is a plasmid which effects high frequency transfer of itself during conjugation and less frequent transfer of the bacterial chromosome itself. A particularly preferred embodiment is to use cloning vectors, referred to as "fosmids" or bacterial artificial chromosome (BAC) vectors. These are derived from *E. coli* f-factor
- 15 which is able to stably integrate large segments of genomic DNA. When integrated with DNA from a mixed uncultured environmental sample, this makes it possible to achieve large genomic fragments in the form of a stable "environmental DNA library."

- 20 Another preferred type of vector for use in the present invention is a cosmid vector. Cosmid vectors were originally designed to clone and propagate large segments of genomic DNA. Cloning into cosmid vectors is described in detail in "Molecular Cloning: A laboratory Manual" (Sambrook et al, 1989).

- 25 The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct RNA synthesis. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I.

Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. Promoter regions can
5 be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed
10 host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g.,
15 the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α -factor, acid phosphatase, or heat shock proteins, among others. The heterologous
20 structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium.

The cloning strategy permits expression via both vector driven and
25 endogenous promoters; vector promotion may be important with expression of genes whose endogenous promoter will not function in *E. coli*.

The DNA isolated or derived from microorganisms can preferably be inserted into a vector or a plasmid prior to probing for selected DNA. Such

vectors or plasmids are preferably those containing expression regulatory sequences, including promoters, enhancers and the like. Such polynucleotides can be part of a vector and/or a composition and still be isolated, in that such vector or composition is not part of its natural environment. Particularly preferred
5 phage or plasmid and methods for introduction and packaging into them are described in detail in the protocol set forth herein.

The selection of the cloning vector depends upon the approach taken, for example, the vector can be any cloning vector with an adequate capacity for
10 multiply repeated copies of a sequence, or multiple sequences that can be successfully transformed and selected in a host cell. One example of such a vector is described in "Polycos vectors: a system for packaging filamentous phage and phagemid vectors using lambda phage packaging extracts" (Alting-Mecs and Short, 1993). Propagation/maintenance can be by an antibiotic resistance carried
15 by the cloning vector. After a period of growth, the naturally abbreviated molecules are recovered and identified by size fractionation on a gel or column, or amplified directly. The cloning vector utilized may contain a selectable gene that is disrupted by the insertion of the lengthy construct. As reductive reassortment progresses, the number of repeated units is reduced and the interrupted gene is
20 again expressed and hence selection for the processed construct can be applied. The vector may be an expression/selection vector which will allow for the selection of an expressed product possessing desirable biological properties. The insert may be positioned downstream of a functional promoter and the desirable property screened by appropriate means.

25

In vivo reassortment is focused on "inter-molecular" processes collectively referred to as "recombination" which in bacteria, is generally viewed as a "RecA-dependent" phenomenon. The present invention can rely on recombination processes of a host cell to recombine and re-assort sequences, or the cells' ability

to mediate reductive processes to decrease the complexity of quasi-repeated sequences in the cell by deletion. This process of "reductive reassortment" occurs by an "intra-molecular", RecA-independent process.

5 Therefore, in another aspect of the present invention, novel polynucleotides can be generated by the process of reductive reassortment. The method involves the generation of constructs containing consecutive sequences (original encoding sequences), their insertion into an appropriate vector, and their subsequent introduction into an appropriate host cell. The reassortment of the
10 individual molecular identities occurs by combinatorial processes between the consecutive sequences in the construct possessing regions of homology, or between quasi-repeated units. The reassortment process recombines and/or reduces the complexity and extent of the repeated sequences, and results in the production of novel molecular species. Various treatments may be applied to
15 enhance the rate of reassortment. These could include treatment with ultra-violet light, or DNA damaging chemicals, and/or the use of host cell lines displaying enhanced levels of "genetic instability". Thus the reassortment process may involve homologous recombination or the natural property of quasi-repeated sequences to direct their own evolution.

20

Repeated or "quasi-repeated" sequences play a role in genetic instability. In the present invention, "quasi-repeats" are repeats that are not restricted to their original unit structure. Quasi-repeated units can be presented as an array of sequences in a construct; consecutive units of similar sequences. Once ligated,
25 the junctions between the consecutive sequences become essentially invisible and the quasi-repetitive nature of the resulting construct is now continuous at the molecular level. The deletion process the cell performs to reduce the complexity of the resulting construct operates between the quasi-repeated sequences. The quasi-repeated units provide a practically limitless repertoire of templates upon

which slippage events can occur. The constructs containing the quasi-repeats thus effectively provide sufficient molecular elasticity that deletion (and potentially insertion) events can occur virtually anywhere within the quasi-repetitive units.

- 5 When the quasi-repeated sequences are all ligated in the same orientation, for instance head to tail or vice versa, the cell cannot distinguish individual units. Consequently, the reductive process can occur throughout the sequences. In contrast, when for example, the units are presented head to head, rather than head to tail, the inversion delineates the endpoints of the adjacent unit so that deletion
- 10 formation will favor the loss of discrete units. Thus, it is preferable with the present method that the sequences are in the same orientation. Random orientation of quasi-repeated sequences will result in the loss of reassortment efficiency, while consistent orientation of the sequences will offer the highest efficiency. However, while having fewer of the contiguous sequences in the same
- 15 orientation decreases the efficiency, it may still provide sufficient elasticity for the effective recovery of novel molecules. Constructs can be made with the quasi-repeated sequences in the same orientation to allow higher efficiency.

Sequences can be assembled in a head to tail orientation using any of a

20 variety of methods, including the following:

- a) Primers that include a poly-A head and poly-T tail which when made single-stranded would provide orientation can be utilized. This is accomplished by having the first few bases of the primers made from RNA and hence easily removed RNaseH.
- 25 b) Primers that include unique restriction cleavage sites can be utilized. Multiple sites, a battery of unique sequences, and repeated synthesis and ligation steps would be required.
- c) The inner few bases of the primer could be thiolated and an exonuclease used to produce properly tailed molecules.

- The recovery of the re-assorted sequences relies on the identification of cloning vectors with a reduced RI. The re-assorted encoding sequences can then be recovered by amplification. The products are re-cloned and expressed. The
- 5 recovery of cloning vectors with reduced RI can be effected by:
- 1) The use of vectors only stably maintained when the construct is reduced in complexity.
 - 2) The physical recovery of shortened vectors by physical procedures. In this case, the cloning vector would be recovered using standard plasmid
 - 10 isolation procedures and size fractionated on either an agarose gel, or column with a low molecular weight cut off utilizing standard procedures.
 - 3) The recovery of vectors containing interrupted genes which can be selected when insert size decreases.
 - 4) The use of direct selection techniques with an expression vector and the
 - 15 appropriate selection.

Encoding sequences (for example, genes) from related organisms may demonstrate a high degree of homology and encode quite diverse protein products. These types of sequences are particularly useful in the present

20 invention as quasi-repeats. However, while the examples illustrated below demonstrate the reassortment of nearly identical original encoding sequences (quasi-repeats), this process is not limited to such nearly identical repeats.

The following example demonstrates the method of the invention.

25 Encoding nucleic acid sequences (quasi-repeats) derived from three (3) unique species are depicted. Each sequence encodes a protein with a distinct set of properties. Each of the sequences differs by a single or a few base pairs at a unique position in the sequence which are designated "A", "B" and "C". The quasi-repeated sequences are separately or collectively amplified and ligated into

random assemblies such that all possible permutations and combinations are available in the population of ligated molecules. The number of quasi-repeat units can be controlled by the assembly conditions. The average number of quasi-repeated units in a construct is defined as the repetitive index (RI).

5

Once formed, the constructs may, or may not be size fractionated on an agarose gel according to published protocols, inserted into a cloning vector, and transfected into an appropriate host cell. The cells are then propagated and "reductive reassortment" is effected. The rate of the reductive reassortment process may be stimulated by the introduction of DNA damage if desired. Whether the reduction in RI is mediated by deletion formation between repeated sequences by an "intra-molecular" mechanism, or mediated by recombination-like events through "inter-molecular" mechanisms is immaterial. The end result is a reassortment of the molecules into all possible combinations.

15

Optionally, the method comprises the additional step of screening the library members of the shuffled pool to identify individual shuffled library members having the ability to bind or otherwise interact (e.g., such as catalytic antibodies) with a predetermined macromolecule, such as for example a proteinaceous receptor, peptide oligosaccharide, viron, or other predetermined compound or structure.

The displayed polypeptides, antibodies, peptidomimetic antibodies, and variable region sequences that are identified from such libraries can be used for therapeutic, diagnostic, research and related purposes (e.g., catalysts, solutes for increasing osmolarity of an aqueous solution, and the like), and/or can be subjected to one or more additional cycles of shuffling and/or affinity selection. The method can be modified such that the step of selecting for a phenotypic characteristic can be other than of binding affinity for a predetermined molecule

25

(e.g., for catalytic activity, stability oxidation resistance, drug resistance, or detectable phenotype conferred upon a host cell).

The present invention provides a method for generating libraries of
5 displayed antibodies suitable for affinity interactions screening. The method
comprises (1) obtaining first a plurality of selected library members comprising a
displayed antibody and an associated polynucleotide encoding said displayed
antibody, and obtaining said associated polynucleotide encoding for said
10 displayed antibody and obtaining said associated polynucleotides or copies
thereof, wherein said associated polynucleotides comprise a region of
substantially identical variable region framework sequence, and (2) introducing
said polynucleotides into a suitable host cell and growing the cells under
conditions which promote recombination and reductive reassortment resulting in
shuffled polynucleotides. CDR combinations comprised by the shuffled pool are
15 not present in the first plurality of selected library members, said shuffled pool
composing a library of displayed antibodies comprising CDR permutations and
suitable for affinity interaction screening. Optionally, the shuffled pool is
subjected to affinity screening to select shuffled library members which bind to a
predetermined epitope (antigen) and thereby selecting a plurality of selected
20 shuffled library members. Further, the plurality of selectively shuffled library
members can be shuffled and screened iteratively, from 1 to about 1000 cycles or
as desired until library members having a desired binding affinity are obtained.

In another aspect of the invention, it is envisioned that prior to or during
25 recombination or reassortment, polynucleotides generated by the method of the
present invention can be subjected to agents or processes which promote the
introduction of mutations into the original polynucleotides. The introduction of
such mutations would increase the diversity of resulting hybrid polynucleotides
and polypeptides encoded therefrom. The agents or processes which promote

mutagenesis can include, but are not limited to: (+)-CC-1065, or a synthetic analog such as (+)-CC-1065-(N3-Adenine, see Sun and Hurley, 1992); an N-acetylated or deacetylated 4'-fluoro-4-aminobiphenyl adduct capable of inhibiting DNA synthesis (see, for example, van de Poll et al, 1992); or a N-acetylated or deacetylated 4-aminobiphenyl adduct capable of inhibiting DNA synthesis (see also, van de Poll et al, 1992, pp. 751-758); trivalent chromium, a trivalent chromium salt, a polycyclic aromatic hydrocarbon ("PAH") DNA adduct capable of inhibiting DNA replication, such as 7-bromomethyl-benz[*a*]anthracene ("BMA"), tris(2,3-dibromopropyl)phosphate ("Tris-BP"), 1,2-dibromo-3-chloropropane ("DBCP"), 2-bromoacrolein (2BA), benzo[*a*]pyrene-7,8-dihydrodiol-9-10-epoxide ("BPDE"), a platinum(II) halogen salt, N-hydroxy-2-amino-3-methylimidazo[4,5-*f*]-quinoline ("N-hydroxy-IQ"), and N-hydroxy-2-amino-1-methyl-6-phenylimidazo[4,5-*f*]-pyridine ("N-hydroxy-PhIP"). Especially preferred "means for slowing or halting PCR amplification consist of UV light (+)-CC-1065 and (+)-CC-1065-(N3-Adenine). Particularly encompassed means are DNA adducts or polynucleotides comprising the DNA adducts from the polynucleotides or polynucleotides pool, which can be released or removed by a process including heating the solution comprising the polynucleotides prior to further processing.

20

In another aspect the present invention is directed to a method of producing recombinant proteins having biological activity by treating a sample comprising double-stranded template polynucleotides encoding a wild-type protein under conditions according to the present invention which provide for the production of hybrid or re-assorted polynucleotides.

25

The invention also provides the use of polynucleotide shuffling to shuffle a population of viral genes (e.g., capsid proteins, spike glycoproteins, polymerases, and proteases) or viral genomes (e.g., paramyxoviridae,

orthomyxoviridae, herpesviruses, retroviruses, reoviruses and rhinoviruses). In an embodiment, the invention provides a method for shuffling sequences encoding all or portions of immunogenic viral proteins to generate novel combinations of epitopes as well as novel epitopes created by recombination; such shuffled viral proteins may comprise epitopes or combinations of epitopes as well as novel epitopes created by recombination; such shuffled viral proteins may comprise epitopes or combinations of epitopes which are likely to arise in the natural environment as a consequence of viral evolution; (e.g., such as recombination of influenza virus strains).

10

The invention also provides a method suitable for shuffling polynucleotide sequences for generating gene therapy vectors and replication-defective gene therapy constructs, such as may be used for human gene therapy, including but not limited to vaccination vectors for DNA-based vaccination, as well as anti-neoplastic gene therapy and other general therapy formats.

15

In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right-hand direction is the carboxy-terminal direction, in accordance with standard usage and convention. Similarly, unless specified otherwise, the left-hand end of single-stranded polynucleotide sequences is the 5' end; the left-hand direction of double-stranded polynucleotide sequences is referred to as the 5' direction. The direction of 5' to 3' addition of nascent RNA transcripts is referred to as the transcription direction; sequence regions on the DNA strand having the same sequence as the RNA and which are 5' to the 5' end of the RNA transcript are referred to as "upstream sequences"; sequence regions on the DNA strand having the same sequence as the RNA and which are 3' to the 3' end of the coding RNA transcript are referred to as "downstream sequences".

20

25

2.11.1. SATURATION MUTAGENESIS

In one aspect, this invention provides for the use of proprietary codon
5 primers (containing a degenerate N,N,G/T sequence) to introduce point mutations
into a polynucleotide, so as to generate a set of progeny polypeptides in which a
full range of single amino acid substitutions is represented at each amino acid
position. The oligos used are comprised contiguously of a first homologous
sequence, a degenerate N,N,G/T sequence, and preferably but not necessarily a
10 second homologous sequence. The downstream progeny translational products
from the use of such oligos include all possible amino acid changes at each amino
acid site along the polypeptide, because the degeneracy of the N,N,G/T sequence
includes codons for all 20 amino acids.

15 In one aspect, one such degenerate oligo (comprised of one degenerate
N,N,G/T cassette) is used for subjecting each original codon in a parental
polynucleotide template to a full range of codon substitutions. In another aspect,
at least two degenerate N,N,G/T cassettes are used – either in the same oligo or
not, for subjecting at least two original codons in a parental polynucleotide
20 template to a full range of codon substitutions. Thus, more than one N,N,G/T
sequence can be contained in one oligo to introduce amino acid mutations at more
than one site. This plurality of N,N,G/T sequences can be directly contiguous, or
separated by one or more additional nucleotide sequence(s). In another aspect,
oligos serviceable for introducing additions and deletions can be used either alone
25 or in combination with the codons containing an N,N,G/T sequence, to introduce
any combination or permutation of amino acid additions, deletions, and/or
substitutions.

In a particular exemplification, it is possible to simultaneously mutagenize two or more contiguous amino acid positions using an oligo that contains contiguous N,N,G/T triplets, i.e. a degenerate (N,N,G/T)_n sequence.

5 In another aspect, the present invention provides for the use of degenerate cassettes having less degeneracy than the N,N,G/T sequence. For example, it may be desirable in some instances to use (e.g. in an oligo) a degenerate triplet sequence comprised of only one N, where said N can be in the first second or third position of the triplet. Any other bases including any combinations and
10 permutations thereof can be used in the remaining two positions of the triplet. Alternatively, it may be desirable in some instances to use (e.g. in an oligo) a degenerate N,N,N triplet sequence, or an N,N, G/C triplet sequence.

 It is appreciated, however, that the use of a degenerate triplet (such as
15 N,N,G/T or an N,N, G/C triplet sequence) as disclosed in the instant invention is advantageous for several reasons. In one aspect, this invention provides a means to systematically and fairly easily generate the substitution of the full range of possible amino acids (for a total of 20 amino acids) into each and every amino acid position in a polypeptide. Thus, for a 100 amino acid polypeptide, the instant
20 invention provides a way to systematically and fairly easily generate 2000 distinct species (i.e. 20 possible amino acids per position X 100 amino acid positions). It is appreciated that there is provided, through the use of an oligo containing a degenerate N,N,G/T or an N,N, G/C triplet sequence, 32 individual sequences that code for 20 possible amino acids. Thus, in a reaction vessel in which a parental
25 polynucleotide sequence is subjected to saturation mutagenesis using one such oligo, there are generated 32 distinct progeny polynucleotides encoding 20 distinct polypeptides. In contrast, the use of a non-degenerate oligo in site-directed mutagenesis leads to only one progeny polypeptide product per reaction vessel.

This invention also provides for the use of nondegenerate oligos, which can optionally be used in combination with degenerate primers disclosed. It is appreciated that in some situations, it is advantageous to use nondegenerate oligos to generate specific point mutations in a working polynucleotide. This provides a means to generate specific silent point mutations, point mutations leading to corresponding amino acid changes, and point mutations that cause the generation of stop codons and the corresponding expression of polypeptide fragments.

Thus, in a preferred embodiment of this invention, each saturation mutagenesis reaction vessel contains polynucleotides encoding at least 20 progeny polypeptide molecules such that all 20 amino acids are represented at the one specific amino acid position corresponding to the codon position mutagenized in the parental polynucleotide. The 32-fold degenerate progeny polypeptides generated from each saturation mutagenesis reaction vessel can be subjected to clonal amplification (e.g. cloned into a suitable *E. coli* host using an expression vector) and subjected to expression screening. When an individual progeny polypeptide is identified by screening to display a favorable change in property (when compared to the parental polypeptide), it can be sequenced to identify the correspondingly favorable amino acid substitution contained therein.

It is appreciated that upon mutagenizing each and every amino acid position in a parental polypeptide using saturation mutagenesis as disclosed herein, favorable amino acid changes may be identified at more than one amino acid position. One or more new progeny molecules can be generated that contain a combination of all or part of these favorable amino acid substitutions. For example, if 2 specific favorable amino acid changes are identified in each of 3 amino acid positions in a polypeptide, the permutations include 3 possibilities at each position (no change from the original amino acid, and each of two favorable

changes) and 3 positions. Thus, there are $3 \times 3 \times 3$ or 27 total possibilities, including 7 that were previously examined - 6 single point mutations (i.e. 2 at each of three positions) and no change at any position.

5 In yet another aspect, site-saturation mutagenesis can be used together with shuffling, chimerization, recombination and other mutagenizing processes, along with screening. This invention provides for the use of any mutagenizing process(es), including saturation mutagenesis, in an iterative manner. In one exemplification, the iterative use of any mutagenizing process(es) is used in
10 combination with screening.

 Thus, in a non-limiting exemplification, this invention provides for the use of saturation mutagenesis in combination with additional mutagenization processes, such as process where two or more related polynucleotides are
15 introduced into a suitable host cell such that a hybrid polynucleotide is generated by recombination and reductive reassortment.

 In addition to performing mutagenesis along the entire sequence of a gene, the instant invention provides that mutagenesis can be use to replace each of any
20 number of bases in a polynucleotide sequence, wherein the number of bases to be mutagenized is preferably every integer from 15 to 100,000. Thus, instead of mutagenizing every position along a molecule, one can subject every a discrete number of bases (preferably a subset totaling from 15 to 100,000) to mutagenesis. Preferably, a separate nucleotide is used for mutagenizing each position or group
25 of positions along a polynucleotide sequence. A group of 3 positions to be mutagenized may be a codon. The mutations are preferably introduced using a mutagenic primer, containing a heterologous cassette, also referred to as a mutagenic cassette. Preferred cassettes can have from 1 to 500 bases. Each nucleotide position in such heterologous cassettes be N, A, C, G, T, A/C, A/G,

A/T, C/G, C/T, G/T, C/G/T, A/G/T, A/C/T, A/C/G, or E, where E is any base that is not A, C, G, or T (E can be referred to as a designer oligo). The tables below show exemplary tri-nucleotide cassettes (there are over 3000 possibilities in addition to N,N,G/T and N,N,N and N,N,A/C).

5

In a general sense, saturation mutagenesis is comprised of mutagenizing a complete set of mutagenic cassettes (wherein each cassette is preferably 1-500 bases in length) in defined polynucleotide sequence to be mutagenized (wherein the sequence to be mutagenized is preferably from 15 to 100,000 bases in length).
10 Thusly, a group of mutations (ranging from 1 to 100 mutations) is introduced into each cassette to be mutagenized. A grouping of mutations to be introduced into one cassette can be different or the same from a second grouping of mutations to be introduced into a second cassette during the application of one round of saturation mutagenesis. Such groupings are exemplified by deletions, additions,
15 groupings of particular codons, and groupings of particular nucleotide cassettes.

Defined sequences to be mutagenized (see Fig. 20) include preferably a whole gene, pathway, cDNA, an entire open reading frame (ORF), and intire promoter, enhancer, repressor/transactivator, origin of replication, intron, operator,
20 or any polynucleotide functional group. Generally, a preferred "defined sequences" for this purpose may be any polynucleotide that a 15 base-polynucleotide sequence, and polynucleotide sequences of lengths between 15 bases and 15,000 bases (this invention specifically names every integer in between). Considerations in choosing groupings of codons include types of
25 amino acids encoded by a degenerate mutagenic cassette.

In a particularly preferred exemplification a grouping of mutations that can be introduced into a mutagenic cassette (see Tables 1-85), this invention specifically provides for degenerate codon substitutions (using degenerate oligos)

that code for 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, and 20 amino acids at each position, and a library of polypeptides encoded thereby.

SUMMARY OF TABLES 1-85

These tables show preferred, but non-limiting, examples of 3-base long mutagenic cassettes that are non-stochastic and degenerate.

Table#	triplet sequence	Site 1	Site 2	Site 3
1.	N,N,G/T	N	N	G/T
2.	N,N,G/C	N	N	G/C
3.	N,N,G/A	N	N	G/A
4.	N,N,A/C	N	N	A/C
5.	N,N,A/T	N	N	A/T
6.	N,N,C/T	N	N	C/T
7.	N,N,N	N	N	N
8.	N,N,G	N	N	G
9.	N,N,A	N	N	A
10.	N,N,C	N	N	C
11.	N,N,T	N	N	T
12.	N,N,C/G/T	N	N	C/G/T
13.	N,N,A/G/T	N	N	A/G/T
14.	N,N,A/C/T	N	N	A/C/T
15.	N,N,A/C/G	N	N	A/C/G
16.	N,A,A	N	A	A
17.	N,A,C	N	A	C
18.	N,A,G	N	A	G
19.	N,A,T	N	A	T
20.	N,C,A	N	C	A
21.	N,C,C	N	C	C
22.	N,C,G	N	C	G
23.	N,C,T	N	C	T
24.	N,G,A	N	G	A
25.	N,G,C	N	G	C
26.	N,G,G	N	G	G
27.	N,G,T	N	G	T
28.	N,T,A	N	T	A
29.	N,T,C	N	T	C
30.	N,T,G	N	T	G
31.	N,T,T	N	T	T
32.	N,A/C,A	N	A/C	A
33.	N,A/G,A	N	A/G	A
34.	N,A/T,A	N	A/T	A
35.	N,C/G,A	N	C/G	A
36.	N,C/T,A	N	C/T	A
37.	N,T/G,A	N	T/G	A
38.	N,C/G/T,A	N	C/G/T	A
39.	N,A/G/T,A	N	A/G/T	A
40.	N,A/C/T,A	N	A/C/T	A
41.	N,A/C/G,A	N	A/C/G	A
42.	A,N,N	A	N	N
43.	C,N,N	C	N	N
44.	G,N,N	G	N	N
45.	T,N,N	T	N	N
46.	A/C,N,N	A/C	N	N

Table#	triplet sequence	Site 1	Site 2	Site 3
47.	A/G,N,N	A/G	N	N
48.	A/T,N,N	A/T	N	N
49.	C/G,N,N	C/G	N	N
50.	C/T,N,N	C/T	N	N
51.	G/T,N,N	G/T	N	N
52.	N,A,N	N	A	N
53.	N,C,N	N	C	N
54.	N,G,N	N	G	N
55.	N,T,N	N	T	N
56.	N,A/C,N	N	A/C	N
57.	N,A/G,N	N	A/G	N
58.	N,A/T,N	N	A/T	N
59.	N,C/G,N	N	C/G	N
60.	N,C/T,N	N	C/T	N
61.	N,G/T,N	N	G/T	N
62.	N,A/C/G,N	N	A/C/G	N
63.	N,A/C/T,N	N	A/C/T	N
64.	N,A/G/T,N	N	A/G/T	N
65.	N,C/G/T,N	N	C/G/T	N
66.	C,C,N	C	C	N
67.	G,G,N	G	G	N
68.	G,C,N	G	C	N
69.	G,T,N	G	T	N
70.	C,G,N	C	G	N
71.	C,T,N	C	T	N
72.	T,C,N	T	C	N
73.	A,C,N	A	C	N
74.	G,A,N	G	A	N
75.	A,T,N	A	T	N
76.	C,A,N	C	A	N
77.	T,T,N	T	T	N
78.	A,A,N	A	A	N
79.	T,A,N	T	A	N
80.	T,G,N	T	G	N
81.	A,G,N	A	G	N
82.	G/C,G,N	G/C	G	N
83.	G/C,C,N	G/C	C	N
84.	G/C,A,N	G/C	A	N
85.	G/C,T,N	G/C	T	N

TABLE 1. Mutagenic Cassette: N, N, G/T

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	2	NONPOLAR (NPL)	15
GGC	NO				
GGA	NO				
GGG	YES				
GCT	YES	ALANINE	2		
GCC	NO				
GCA	NO				
GCG	YES				
GTT	YES	VALINE	2		
GTC	NO				
GTA	NO				
GTG	YES				
TTA	NO	LEUCINE	3		
TTG	YES				
CTT	YES				
CTC	NO				
CTA	NO				
CTG	YES	ISOLEUCINE	1		
ATT	YES				
ATC	NO				
ATA	NO	METHIONINE	1		
ATG	YES				
TTT	YES	PHENYLALANINE	1		
TTC	NO				
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	2		
CCC	NO				
CCA	NO				
CCG	YES				
TCT	YES	SERINE	3	POLAR NONIONIZABLE (POL)	9
TCC	NO				
TCA	NO				
TCG	YES				
AGT	YES				
AGC	NO				
TGT	YES	CYSTEINE	1		
TGC	NO				
AAT	YES	ASPARAGINE	1		
AAC	NO				
CAA	NO	GLUTAMINE	1		
CAG	YES				
TAT	YES	TYROSINE	1		
TAC	NO				
ACT	YES	THREONINE	2		
ACC	NO				
ACA	NO				
ACG	YES				
GAT	YES	ASPARTIC ACID	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	2
GAC	NO				
GAA	NO	GLUTAMIC ACID	1		
GAG	YES				
AAA	NO	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	5
AAG	YES				
CGT	YES	ARGININE	3		
CGC	NO				
CGA	NO				
CGG	YES				
AGA	NO				
AGG	YES				
CAT	YES	HISTIDINE	1		
CAC	NO				
TAA	NO	STOP CODON	1	STOP SIGNAL (STP)	1
TAG	YES				
TGA	NO				
64	32	20 Amino Acids Are Represented		NPL: 15: POL: 9: NEG: 2: POS: 5: STP: 1	

TABLE 2. Mutagenic Cassette: N, N, G/C

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	NO	GLYCINE	2	NONPOLAR (NPL)	15
GGC	YES				
GGA	NO				
GGG	YES				
GCT	NO	ALANINE	2		
GCC	YES				
GCA	NO				
GCG	YES				
GTT	NO	VALINE	2		
GTC	YES				
GTA	NO				
GTG	YES				
TTA	NO	LEUCINE	3		
TTG	YES				
CTT	NO				
CTC	YES				
CTA	NO				
CTG	YES	ISOLEUCINE	1		
ATT	NO				
ATC	YES				
ATA	NO				
ATG	YES	METHIONINE	1		
TTT	NO	PHENYLALANINE	1		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
CCT	NO	PROLINE	2		
CCC	YES				
CCA	NO				
CCG	YES				
TCT	NO	SERINE	3	POLAR NONIONIZABLE (POL)	9
TCC	YES				
TCA	NO				
TCG	YES				
AGT	NO				
AGC	YES	CYSTEINE	1		
TGT	NO				
TGC	YES				
AAT	NO	ASPARAGINE	1		
AAC	YES				
CAA	NO	GLUTAMINE	1		
CAG	YES				
TAT	NO	TYROSINE	1		
TAC	YES				
ACT	NO	THREONINE	2		
ACC	YES				
ACA	NO				
ACG	YES				
GAT	NO	ASPARTIC ACID	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	2
GAC	YES				
GAA	NO	GLUTAMIC ACID	1		
GAG	YES				
AAA	NO	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	5
AAG	YES				
CGT	NO	ARGININE	3		
CGC	YES				
CGA	NO				
CGG	YES				
AGA	NO				
AGG	YES	HISTIDINE	1		
CAT	NO				
CAC	YES				
TAA	NO	STOP CODON	1	STOP SIGNAL (STP)	1
TAG	YES				
TGA	NO				
64	32	20 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 15: 9: 2: 5: 1	

TABLE 3. Mutagenic Cassette: N, N, G/A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	NO	GLYCINE	2	NONPOLAR (NPL)	15
GGC	NO				
GGA	YES				
GGG	YES				
GCT	NO	ALANINE	2		
GCC	NO				
GCA	YES				
GCG	YES				
GTT	NO	VALINE	2		
GTC	NO				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	4		
TTG	YES				
CTT	NO				
CTC	NO				
CTA	YES				
CTG	YES				
ATT	NO	ISOLEUCINE	1		
ATC	NO				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	NO	PHENYLALANINE	0		
TTC	NO				
TGG	YES	TRYPTOPHAN	1		
CCT	NO	PROLINE	2		
CCC	NO				
CCA	YES				
CCG	YES				
TCT	NO	SERINE	2	POLAR NONIONIZABLE (POL)	6
TCC	NO				
TCA	YES				
TCG	YES				
AGT	NO				
AGC	NO	CYSTEINE	0		
TGT	NO				
TGC	NO				
AAT	NO	ASPARAGINE	0		
AAC	NO				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	NO	TYROSINE	0		
TAC	NO				
ACT	NO	THREONINE	2		
ACC	NO				
ACA	YES				
ACG	YES				
GAT	NO	ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	2
GAC	NO	GLUTAMIC ACID	2		
GAA	YES				
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
AAG	YES				
CGT	NO	ARGININE	4		
CGC	NO				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES	HISTIDINE	0		
CAT	NO				
CAC	NO				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
64	32	14 Amino Acids Are Represented		NPL: 15: POL: 6: NEG: 2: POS: 6: STP: 3	

TABLE 4. Mutagenic Cassette: N, N, A/C

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	NO	GLYCINE	2	NONPOLAR (NPL)	14
GGC	YES				
GGA	YES				
GGG	NO				
GCT	NO	ALANINE	2		
GCC	YES				
GCA	YES				
GCG	NO				
GTT	NO	VALINE	2		
GTC	YES				
GTA	YES				
GTG	NO				
TTA	YES	LEUCINE	3		
TTG	NO				
CTT	NO				
CTC	YES				
CTA	YES				
CTG	NO				
ATT	NO	ISOLEUCINE	2		
ATC	YES				
ATA	YES				
ATG	NO	METHIONINE	0		
TTT	NO	PHENYLALANINE	1		
TTC	YES				
TGG	NO	TRYPTOPHAN	0		
CCT	NO	PROLINE	2		
CCC	YES				
CCA	YES				
CCG	NO				
TCT	NO	SERINE	3	POLAR NONIONIZABLE (POL)	9
TCC	YES				
TCA	YES				
TCG	NO				
AGT	NO				
AGC	YES				
TGT	NO	CYSTEINE	1		
TGC	YES				
AAT	NO	ASPARAGINE	1		
AAC	YES				
CAA	YES	GLUTAMINE	1		
CAG	NO				
TAT	NO	TYROSINE	1		
TAC	YES				
ACT	NO	THREONINE	2		
ACC	YES				
ACA	YES				
ACG	NO				
GAT	NO	ASPARTIC ACID	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	2
GAC	YES	GLUTAMIC ACID	1		
GAA	YES				
GAG	NO				
AAA	YES	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	5
AAG	NO				
CGT	NO	ARGININE	3		
CGC	YES				
CGA	YES				
CGG	NO				
AGA	YES				
AGG	NO				
CAT	NO	HISTIDINE	1		
CAC	YES				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	NO				
TGA	YES				
64	32	18 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 14: 9: 2: 5: 2	

TABLE 5. Mutagenic Cassette: N, N, A/T

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	2	NONPOLAR (NPL)	14
GGC	NO				
GGA	YES				
GGG	NO				
GCT	YES	ALANINE	2		
GCC	NO				
GCA	YES				
GCG	NO				
GTT	YES	VALINE	2		
GTC	NO				
GTA	YES				
GTG	NO				
TTA	YES	LEUCINE	3		
TTG	NO				
CTT	YES				
CTC	NO				
CTA	YES				
CTG	NO				
ATT	YES	ISOLEUCINE	2		
ATC	NO				
ATA	YES				
ATG	NO	METHIONINE	0		
TTT	YES	PHENYLALANINE	1		
TTC	NO				
TGG	NO	TRYPTOPHAN	0		
CCT	YES	PROLINE	2		
CCC	NO				
CCA	YES				
CCG	NO				
TCT	YES	SERINE	3	POLAR NONIONIZABLE (POL)	9
TCC	NO				
TCA	YES				
TCG	NO				
AGT	YES				
AGC	NO				
TGT	YES	CYSTEINE	1		
TGC	NO				
AAT	YES	ASPARAGINE	1		
AAC	NO				
CAA	YES	GLUTAMINE	1		
CAG	NO				
TAT	YES	TYROSINE	1		
TAC	NO				
ACT	YES	THREONINE	2		
ACC	NO				
ACA	YES				
ACG	NO				
GAT	YES	ASPARTIC ACID	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	2
GAC	NO				
GAA	YES	GLUTAMIC ACID	1		
GAG	NO				
AAA	YES	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	5
AAG	NO				
CGT	YES	ARGININE	3		
CGC	NO				
CGA	YES				
CGG	NO				
AGA	YES				
AGG	NO				
CAT	YES	HISTIDINE	1		
CAC	NO				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	NO				
TGA	YES				
64	32	18 Amino Acids Are Represented		NPL: 14: POL: 9: NEG: 2: POS: 5: STP = 2	

TABLE 6. Mutagenic Cassette: N, N, C/T

CODON	Represented	AMINO ACID (Frequency)	CATEGORY (Frequency)	
GGT	YES	GLYCINE 2	NONPOLAR (NPL) 14	
GGC	YES			
GGA	NO			
GGG	NO			
GCT	YES	ALANINE 2		
GCC	YES			
GCA	NO			
GCG	NO			
GTT	YES	VALINE 2		
GTC	YES			
GTA	NO			
GTG	NO			
TTA	NO	LEUCINE 2		
TTG	NO			
CTT	YES			
CTC	YES			
CTA	NO			
CTG	NO			
ATT	YES	ISOLEUCINE 2		
ATC	YES			
ATA	NO			
ATG	NO	METHIONINE 0		
TTT	YES	PHENYLALANINE 2		
TTC	YES			
TGG	NO	TRYPTOPHAN 0		
CCT	YES	PROLINE 2		
CCC	YES			
CCA	NO			
CCG	NO			
TCT	YES	SERINE 4	POLAR NONIONIZABLE (POL) 12	
TCC	YES			
TCA	NO			
TCG	NO			
AGT	YES			
AGC	YES			
TGT	YES			CYSTEINE 2
TGC	YES			
AAT	YES			ASPARAGINE 2
AAC	YES			GLUTAMINE 0
CAA	NO			
CAG	NO			TYROSINE 2
TAT	YES			
TAC	YES			THREONINE 2
ACT	YES			
ACC	YES			
ACA	NO			
ACG	NO			
GAT	YES	ASPARTIC ACID 2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG) 2	
GAC	YES	GLUTAMIC ACID 0		
GAA	NO			
GAG	NO			
AAA	NO	LYSINE 0	IONIZABLE: BASIC POSITIVE CHARGE (POS) 4	
AAG	NO	ARGININE 2		
CGT	YES			
CGC	YES			
CGA	NO			
CGG	NO			
AGA	NO	HISTIDINE 2		
AGG	NO			
CAT	YES			
CAC	YES			
TAA	NO	STOP CODON 0	STOP SIGNAL (STP) 0	
TAG	NO			
TGA	NO			
64	32	15 Amino Acids Are Represented	NPL: 14: POL: 12: NEG: 2: POS: 4: STP: 0	

TABLE 7. Mutagenic Cassette: N, N, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	29
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	6	POLAR NONIONIZABLE (POL)	18
TCC	YES				
TCA	YES				
TCG	YES				
AGT	YES				
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10
AAG	YES				
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
64	64	20 Amino Acids Are Represented		NPL: 29: POL: 18: NEG: 4: POS: 10: STP: 3	

TABLE 8. Mutagenic Cassette: N, N, G

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	NO	GLYCINE	1	NONPOLAR (NPL)	8
GGC	NO				
GGA	NO				
GGG	YES				
GCT	NO	ALANINE	1		
GCC	NO				
GCA	NO				
GCG	YES				
GTT	NO	VALINE	1		
GTC	NO				
GTA	NO				
GTG	YES				
TTA	NO	LEUCINE	2		
TTG	YES				
CTT	NO				
CTC	NO				
CTA	NO				
CTG	YES				
ATT	NO	ISOLEUCINE	0		
ATC	NO				
ATA	NO				
ATG	YES	METHIONINE	1		
TTT	NO	PHENYLALANINE	0		
TTC	NO				
TGG	YES	TRYPTOPHAN	1		
CCT	NO	PROLINE	1		
CCC	NO				
CCA	NO				
CCG	YES				
TCT	NO	SERINE	1	POLAR NONIONIZABLE (POL)	3
TCC	NO				
TCA	NO				
TCG	YES				
AGT	NO	CYSTEINE	0		
AGC	NO				
TGT	NO				
TGC	NO				
AAT	NO	ASPARAGINE	0		
AAC	NO				
CAA	NO	GLUTAMINE	1		
CAG	YES				
TAT	NO	TYROSINE	0		
TAC	NO				
ACT	NO	THREONINE	1		
ACC	NO				
ACA	NO				
ACG	YES				
GAT	NO	ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
GAC	NO	GLUTAMIC ACID	1		
GAA	NO				
GAG	YES				
AAA	NO	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	3
AAG	YES	ARGININE	2		
CGT	NO				
CGC	NO				
CGA	NO				
CGG	YES				
AGA	NO	HISTIDINE	0		
AGG	YES				
CAT	NO				
CAC	NO				
TAA	NO	STOP CODON	1	STOP SIGNAL (STP)	1
TAG	YES				
TGA	NO				
64	16	13 Amino Acids Are Represented		NPL: 8: POL: 3: NEG: 1: POS: 3: STP: 1	

TABLE 9. Mutagenic Cassette: N, N, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	NO	GLYCINE	1	NONPOLAR (NPL)	7
GGC	NO				
GGA	YES				
GGG	NO				
GCT	NO	ALANINE	1		
GCC	NO				
GCA	YES				
GCG	NO				
GTT	NO	VALINE	1		
GTC	NO				
GTA	YES				
GTG	NO				
TTA	YES	LEUCINE	2		
TTG	NO				
CTT	NO				
CTC	NO				
CTA	YES				
CTG	NO	ISOLEUCINE	1		
ATT	NO				
ATC	NO				
ATA	YES	METHIONINE	0		
ATG	NO				
TTT	NO	PHENYLALANINE	0		
TTC	NO				
TGG	NO	TRYPTOPHAN	0		
CCT	NO				
CCC	NO	PROLINE	1		
CCA	YES				
CCG	NO				
TCT	NO	SERINE	1		
TCC	NO				
TCA	YES				
TCG	NO				
AGT	NO				
AGC	NO	CYSTEINE	0		
TGT	NO				
TGC	NO				
AAT	NO	ASPARAGINE	0		
AAC	NO				
CAA	YES	GLUTAMINE	1		
CAG	NO				
TAT	NO	TYROSINE	0		
TAC	NO				
ACT	NO	THREONINE	1		
ACC	NO				
ACA	YES				
ACG	NO				
GAT	NO	ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
GAC	NO				
GAA	YES	GLUTAMIC ACID	1		
GAG	NO				
AAA	YES	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	3
AAG	NO				
CGT	NO	ARGININE	2		
CGC	NO				
CGA	YES				
CGG	NO				
AGA	YES				
AGG	NO	HISTIDINE	0		
CAT	NO				
CAC	NO				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	NO				
TGA	YES				
64	16	12 Amino Acids Are Represented		NPL: 7; POL: 3; NEG: 1; POS: 3; STP: 2	

TABLE 10. Mutagenic Cassette: N, N, C

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)		
GGT	NO	GLYCINE	1	NONPOLAR (NPL)	7		
GGC	YES						
GGA	NO						
GGG	NO						
GCT	NO	ALANINE	1				
GCC	YES						
GCA	NO						
GCG	NO						
GTT	NO	VALINE	1				
GTC	YES						
GTA	NO						
GTG	NO						
TTA	NO	LEUCINE	1				
TTG	NO						
CTT	NO						
CTC	YES						
CTA	NO	ISOLEUCINE	1				
CTG	NO						
ATT	NO						
ATC	YES						
ATA	NO	METHIONINE	0				
ATG	NO						
TTT	NO						
TTC	YES						
TGG	NO	TRYPTOPHAN	0				
CCT	NO	PROLINE	1				
CCC	YES						
CCA	NO						
CCG	NO						
TCT	NO	SERINE	2	POLAR NONIONIZABLE (POL)	6		
TCC	YES						
TCA	NO						
TCG	NO						
AGT	NO	CYSTEINE	1				
AGC	YES						
TGT	NO						
TGC	YES						
AAT	NO	ASPARAGINE	1				
AAC	YES						
CAA	NO						
CAG	NO						
TAT	NO	TYROSINE	1				
TAC	YES						
ACT	NO						
ACC	YES						
ACA	NO	THREONINE	1				
ACG	NO						
GAT	NO						
GAC	YES	ASPARTIC ACID	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1		
GAA	NO	GLUTAMIC ACID	0				
GAG	NO						
AAA	NO	LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2		
AAG	NO						
CGT	NO						
CGC	YES						
CGA	NO	ARGININE	1				
CGG	NO						
AGA	NO						
AGG	NO						
CAT	NO	HISTIDINE	1				
CAC	YES						
TAA	NO	STOP CODON	0	STOP SIGNAL (STP)	0		
TAG	NO						
TGA	NO						
64	16	15 Amino Acids Are Represented		NPL: 7: POL: 6: NEG: 1: POS: 2: STP: 0			

TABLE 11. Mutagenic Cassette: N, N, T

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT		YES	GLYCINE	1	NONPOLAR (NPL)	7
GGC		NO				
GGA		NO				
GGG		NO				
GCT		YES	ALANINE	1		
GCC		NO				
GCA		NO				
GCG		NO				
GTT		YES	VALINE	1		
GTC		NO				
GTA		NO				
GTG		NO				
TTA		NO	LEUCINE	1		
TTG		NO				
CTT		YES				
CTC		NO				
CTA		NO				
CTG		NO				
ATT		YES	ISOLEUCINE	1		
ATC		NO				
ATA		NO				
ATG		NO	METHIONINE	0		
TTT		YES	PHENYLALANINE	1		
TTC		NO				
TGG		NO	TRYPTOPHAN	0		
CCT		YES	PROLINE	1		
CCC		NO				
CCA		NO				
CCG		NO				
TCT		YES	SERINE	2	POLAR NONIONIZABLE (POL)	6
TCC		NO				
TCA		NO				
TCG		NO				
AGT		YES				
AGC		NO				
TGT		YES	CYSTEINE	1		
TGC		NO				
AAT		YES	ASPARAGINE	1		
AAC		NO				
CAA		NO	GLUTAMINE	0		
CAG		NO				
TAT		YES	TYROSINE	1		
TAC		NO				
ACT		YES	THREONINE	1		
ACC		NO				
ACA		NO				
ACG		NO				
GAT		YES	ASPARTIC ACID	1		
GAC		NO				
GAA		NO	GLUTAMIC ACID	0		
GAG		NO				
AAA		NO	LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
AAG		NO				
CGT		YES	ARGININE	1		
CGC		NO				
CGA		NO				
CGG		NO				
AGA		NO				
AGG		NO	HISTIDINE	1		
CAT		YES				
CAC		NO				
TAA		NO	STOP CODON	0	STOP SIGNAL (STP)	0
TAG		NO				
TGA		NO				
64		16	15 Amino Acids Are Represented		NPL: 7: POL: 6: NEG: 1: POS: 2: STP: 0	

TABLE 12. Mutagenic Cassette: N, N, C/G/T

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	3	NONPOLAR (NPL)	22
GGC	YES				
GGA	NO				
GGG	YES				
GCT	YES	ALANINE	3		
GCC	YES				
GCA	NO				
GCG	YES				
GTT	YES	VALINE	3		
GTC	YES				
GTA	NO				
GTG	YES				
TTA	NO	LEUCINE	4		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	NO				
CTG	YES				
ATT	YES	ISOLEUCINE	2		
ATC	YES				
ATA	NO				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	3		
CCC	YES				
CCA	NO				
CCG	YES				
TCT	YES	SERINE	5	POLAR NONIONIZABLE (POL)	15
TCC	YES				
TCA	NO				
TCG	YES				
AGT	YES				
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	NO	GLUTAMINE	1		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	3		
ACC	YES				
ACA	NO				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	3
GAC	YES	GLUTAMIC ACID	1		
GAA	NO				
GAG	YES				
AAA	NO	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	7
AAG	YES				
CGT	YES	ARGININE	4		
CGC	YES				
CGA	NO				
CGG	YES				
AGA	NO				
AGG	YES				
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	NO	STOP CODON	1	STOP SIGNAL (STP)	1
TAG	YES				
TGA	NO				
64	48	20 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 22: 15: 3: 7: 1	

TABLE 13. Mutagenic Cassette: N, N, A/G/T

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	3	NONPOLAR (NPL)	22
GGC	NO				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	3		
GCC	NO				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	3		
GTC	NO				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	5		
TTG	YES				
CTT	YES				
CTC	NO				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	2		
ATC	NO				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	1		
TTC	NO	TRYPTOPHAN	1		
TGG	YES				
CCT	YES				
CCC	NO	PROLINE	3		
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	12
TCC	NO				
TCA	YES				
TCG	YES				
AGT	YES				
AGC	NO				
TGT	YES	CYSTEINE	1		
TGC	NO				
AAT	YES	ASPARAGINE	1		
AAC	NO	GLUTAMINE	2		
CAA	YES				
CAG	YES	TYROSINE	1		
TAT	YES				
TAC	NO	THREONINE	3		
ACT	YES				
ACC	NO				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	3
GAC	NO	GLUTAMIC ACID	2		
GAA	YES				
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	8
AAG	YES	ARGININE	5		
CGT	YES				
CGC	NO				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
CAT	YES	HISTIDINE	1		
CAC	NO				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
64	48	20 Amino Acids Are Represented		NPL: 22: POL: 12: NEG: 3: POS: 8: STP: 3	

TABLE 14. Mutagenic Cassette: N, N, A/C/T

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	3	NONPOLAR (NPL)	21
GGC	YES				
GGA	YES				
GGG	NO				
GCT	YES	ALANINE	3		
GCC	YES				
GCA	YES				
GCG	NO				
GTT	YES	VALINE	3		
GTC	YES				
GTA	YES				
GTG	NO				
TTA	YES	LEUCINE	4		
TTG	NO				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	NO				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	NO	METHIONINE	0		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	NO	TRYPTOPHAN	0		
CCT	YES	PROLINE	3		
CCC	YES				
CCA	YES				
CCG	NO				
TCT	YES	SERINE	5	POLAR NONIONIZABLE (POL)	15
TCC	YES				
TCA	YES				
TCG	NO				
AGT	YES				
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	1		
CAG	NO				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	3		
ACC	YES				
ACA	YES				
ACG	NO				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	3
GAC	YES				
GAA	YES	GLUTAMIC ACID	1		
GAG	NO				
AAA	YES	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	7
AAG	NO				
CGT	YES	ARGININE	4		
CGC	YES				
CGA	YES				
CGG	NO				
AGA	YES				
AGG	NO				
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	NO				
TGA	YES				
64	48	18 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 21: 15: 3: 7: 2	

TABLE 15. Mutagenic Cassette: N, N, A/C/G

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	NO	GLYCINE	3	NONPOLAR (NPL)	22
GGC	YES				
GGA	YES				
GGG	YES				
GCT	NO	ALANINE	3		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	NO	VALINE	3		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	5		
TTG	YES				
CTT	NO				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	NO	ISOLEUCINE	2		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	NO	PHENYLALANINE	1		
TTC	YES	TRYPTOPHAN	1		
TGG	YES				
CCT	NO	PROLINE	3		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	NO	SERINE	4	POLAR NONIONIZABLE (POL)	12
TCC	YES				
TCA	YES				
TCG	YES				
AGT	NO				
AGC	YES				
TGT	NO	CYSTEINE	1		
TGC	YES				
AAT	NO	ASPARAGINE	1		
AAC	YES	GLUTAMINE	2		
CAA	YES				
CAG	YES				
TAT	NO	TYROSINE	1		
TAC	YES				
ACT	NO	THREONINE	3		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	NO	ASPARTIC ACID	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	3
GAC	YES	GLUTAMIC ACID	2		
GAA	YES				
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	8
AAG	YES				
CGT	NO	ARGININE	5		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
CAT	NO	HISTIDINE	1		
CAC	YES				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
64	48	20 Amino Acids Are Represented		NPL: 22: POL: 12: NEG: 3: POS: 8: STP: 3	

TABLE 16. Mutagenic Cassette: N, A, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	1
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	1		
		TYROSINE	0		
		THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
GAA	YES	ASPARTIC ACID	0		
		GLUTAMIC ACID	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
AAA	YES	LYSINE	1		
		ARGININE	0		
		HISTIDINE	0		
TAA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	4	3 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 0: 1: 1: 1: 1	

TABLE 17. Mutagenic Cassette: N, A, C

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
AAC	YES	ASPARAGINE	1		
		GLUTAMINE	0		
TAC	YES	TYROSINE	1		
		THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
GAC	YES	ASPARTIC ACID	1		
		GLUTAMIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
		LYSINE	0		
		ARGININE	0		
CAC	YES	HISTIDINE	1	STOP SIGNAL (STP)	0
		STOP CODON	0		
TOTAL	4	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 0: 2: 1: 1: 0	

TABLE 18. Mutagenic Cassette: N, A, G

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	1
		CYSTEINE	0		
		ASPARAGINE	0		
CAG	YES	GLUTAMINE	1		
		TYROSINE	0		
		THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
GAG	YES	GLUTAMIC ACID	1		
AAG	YES	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
		ARGININE	0		
		HISTIDINE	0		
TAG	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	4	3 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 0: 1: 1: 1: 1	

TABLE 19. Mutagenic Cassette: N, A, T

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	0
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
			SERINE	0	POLAR NONIONIZABLE (POL)	2
			CYSTEINE	0		
AAT	YES		ASPARAGINE	1		
			GLUTAMINE	0		
TAT	YES		TYROSINE	1		
			THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
GAT	YES		ASPARTIC ACID	1		
			GLUTAMIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
			LYSINE	0		
			ARGININE	0		
CAT	YES		HISTIDINE	1	STOP SIGNAL (STP)	0
			STOP CODON	0		
TOTAL		4	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 0: 2: 1: 1: 0	

TABLE 20. Mutagenic Cassette: N, C, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GCA	YES	GLYCINE	0	NONPOLAR (NPL)	2
		ALANINE	1		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCA	YES	PROLINE	1	POLAR NONIONIZABLE (POL)	2
TCA	YES	SERINE	1		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACA	YES	THREONINE	1		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 2: 2: 0: 0: 0	

TABLE 21. Mutagenic Cassette: N, C, C

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GCC	YES	GLYCINE	0	NONPOLAR (NPL)	2
		ALANINE	1		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCC	YES	PROLINE	1	POLAR NONIONIZABLE (POL)	2
TCC	YES	SERINE	1		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACC	YES	THREONINE	1		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 2: 2: 0: 0: 0	

TABLE 22. Mutagenic Cassette: N, C, G

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	2
GCG	YES	ALANINE	1		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCG	YES	PROLINE	1	POLAR NONIONIZABLE (POL)	2
TCG	YES	SERINE	1		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACG	YES	THREONINE	1		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	4 Amino Acids Are Represented		NPL: 2: POL: 2: NEG: 0: POS: 0: STP: 0	

TABLE 23. Mutagenic Cassette: N, C, T

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	2
GCT	YES	ALANINE	1		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	1	POLAR NONIONIZABLE (POL)	2
TCT	YES	SERINE	1		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	1		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	4 Amino Acids Are Represented		NPL: 2: POL: 2: NEG: 0: POS: 0: STP: 0	

TABLE 24. Mutagenic Cassette: N, G, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGA	YES	GLYCINE	1	NONPOLAR (NPL)	1
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
CGA	YES	LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
AGA	YES	ARGININE	2		
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 1: 0: 0: 2: 1	

TABLE 25. Mutagenic Cassette: N, G, C

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGC	YES	GLYCINE	1	NONPOLAR (NPL)	1
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
AGC	YES	SERINE	1	POLAR NONIONIZABLE (POL)	2
TGC	YES	CYSTEINE	1		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
CGC	YES	ARGININE	1		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 1: 2: 0: 1: 0	

TABLE 26. Mutagenic Cassette: N, G, G

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGG	YES	GLYCINE	1	NONPOLAR (NPL)	2
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0	POLAR NONIONIZABLE (POL)	0
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
CGG	YES	LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
AGG	YES	ARGININE	2		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	3 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 2: 0: 0: 2: 0	

TABLE 27. Mutagenic Cassette: N, G, T

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	1	NONPOLAR (NPL)	1
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0	POLAR NONIONIZABLE (POL)	2
AGT	YES	SERINE	1		
TGT	YES	CYSTEINE	1		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
CGT	YES	LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
		ARGININE	1		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 1: 2: 0: 1: 0	

TABLE 28. Mutagenic Cassette: N, T, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
GTA	YES	VALINE	1		
TTA	YES	LEUCINE	2		
CTA	YES				
ATA	YES	ISOLEUCINE	1		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	3 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 4: 0: 0: 0: 0	

TABLE 29. Mutagenic Cassette: N, T, C

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
GTC	YES	VALINE	1		
CTC	YES	LEUCINE	1		
ATC	YES	ISOLEUCINE	1		
		METHIONINE	0		
TTC	YES	PHENYLALANINE	1		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 4: 0: 0: 0: 0	

TABLE 30. Mutagenic Cassette: N, T, G

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
GTG	YES	VALINE	1		
TTG	YES	LEUCINE	2		
CTG	YES				
		ISOLEUCINE	0		
ATG	YES	METHIONINE	1		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	3 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 4: 0: 0: 0: 0	

TABLE 31. Mutagenic Cassette: N, T, T

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
GTT	YES	VALINE	1		
CTT	YES	LEUCINE	1		
ATT	YES	ISOLEUCINE	1		
		METHIONINE	0		
TTT	YES	PHENYLALANINE	1		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 4: 0: 0: 0: 0	

TABLE 32. Mutagenic Cassette: N, A/C, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	2
GCA	YES	ALANINE	1		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCA	YES	PROLINE	1	POLAR NONIONIZABLE (POL)	3
TCA	YES	SERINE	1		
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	1		
		TYROSINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
ACA	YES	THREONINE	1		
GAA	YES	ASPARTIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
		GLUTAMIC ACID	1		
AAA	YES	LYSINE	1		
		ARGININE	0	STOP SIGNAL (STP)	1
		HISTIDINE	0		
TAA	YES	STOP CODON	1	NPL: POL: NEG: POS: STP =	
				2: 3: 1: 1: 1	
TOTAL	8	7 Amino Acids Are Represented			

TABLE 33. Mutagenic Cassette: N, A/G, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGA	YES	GLYCINE	1	NONPOLAR (NPL)	1
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0	POLAR NONIONIZABLE (POL)	1
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	1		
		TYROSINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
		THREONINE	0		
GAA	YES	ASPARTIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	3
		GLUTAMIC ACID	1		
AAA	YES	LYSINE	1	STOP SIGNAL (STP)	2
CGA	YES	ARGININE	2		
AGA	YES	HISTIDINE	0		
TAA	YES	STOP CODON	2		
TGA	YES				
TOTAL	8	5 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 1: 1: 1: 3: 2	

TABLE 34. Mutagenic Cassette: N, A/T, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
GTA	YES	VALINE	1		
TTA	YES	LEUCINE	2		
CTA	YES				
ATA	YES	ISOLEUCINE	1		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	1
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	1		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
GAA	YES	GLUTAMIC ACID	1		
AAA	YES	LYSINE	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
		ARGININE	0		
		HISTIDINE	0		
TAA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	8	6 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 4: 1: 1: 1: 1	

TABLE 35. Mutagenic Cassette: N, C/G, A

Multiple Choice: 14, 15, 16, 17					
CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGA	YES	GLYCINE	1	NONPOLAR (NPL)	3
GCA	YES	ALANINE	1		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCA	YES	PROLINE	1	POLAR NONIONIZABLE (POL)	2
TCA	YES	SERINE	1		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACA	YES	THREONINE	1		
		ASPARTIC ACID	0		
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
CGA	YES	ARGININE	2		
AGA	YES				
		HISTIDINE	0	STOP SIGNAL (STP)	1
TGA	YES	STOP CODON	1		
TOTAL	8	6 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 3: 2: 0: 2: 1	

TABLE 36. Mutagenic Cassette: N, C/T, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	6
GCA	YES	ALANINE	1		
GTA	YES	VALINE	1		
TTA	YES	LEUCINE	2		
CTA	YES				
ATA	YES	ISOLEUCINE	1		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCA	YES	PROLINE	1		
TCA	YES	SERINE	1	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACA	YES	THREONINE	1		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	8	7 Amino Acids Are Represented		NPL: 6: POL: 2: NEG: 0: POS: 0: STP: 0	

TABLE 37. Mutagenic Cassette: N, T/G, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGA	YES	GLYCINE	1	NONPOLAR (NPL)	5
		ALANINE	0		
GTA	YES	VALINE	1		
TTA	YES	LEUCINE	2		
CTA	YES				
ATA	YES	ISOLEUCINE	1		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
CGA	YES	ARGININE	2		
AGA	YES				
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	8	5 Amino Acids Are Represented		NPL: 5: POL: 0: NEG: 0: POS: 2: STP: 1	

TABLE 38. Mutagenic Cassette: N, C/G/T, A

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGA	YES		GLYCINE	1	NONPOLAR (NPL)	7
GCA	YES		ALANINE	1		
GTA	YES		VALINE	1		
TTA	YES		LEUCINE	2		
CTA	YES					
ATA	YES		ISOLEUCINE	1		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
CCA	YES		PROLINE	1		
TCA	YES		SERINE	1	POLAR NONIONIZABLE (POL)	2
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
ACA	YES		THREONINE	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			ASPARTIC ACID	0		
			GLUTAMIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
			LYSINE	0		
CGA	YES		ARGININE	2		
AGA	YES				STOP SIGNAL (STP)	1
			HISTIDINE	0		
TGA	YES		STOP CODON	1	NPL: POL: NEG: POS: STP = 7: 2: 0: 2: 1	
TOTAL		12	9 Amino Acids Are Represented			

TABLE 39. Mutagenic Cassette: N, A/G/T, A

TABLE 39. Mutagenic Casette: 19, ACGT, 1, A					
CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGA	YES	GLYCINE	1	NONPOLAR (NPL)	5
		ALANINE	0		
GTA	YES	VALINE	1		
TTA	YES	LEUCINE	2		
CTA	YES				
ATA	YES	ISOLEUCINE	1		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	1
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	1		
		TYROSINE	0		
		THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
		ASPARTIC ACID	0		
GAA	YES	GLUTAMIC ACID	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	3
AAA	YES	LYSINE	1		
CGA	YES	ARGININE	2		
AGA	YES			STOP SIGNAL (STP)	2
		HISTIDINE	0		
TAA	YES	STOP CODON	2	NPL: POL: NEG: POS: STP = 5: 1: 1: 3: 2	
TGA	YES				
TOTAL	12	8 Amino Acids Are Represented			

TABLE 40. Mutagenic Cassette: N, A/C/T, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	6
GCA	YES	ALANINE	1		
GTA	YES	VALINE	1		
TTA	YES	LEUCINE	2		
CTA	YES				
ATA	YES	ISOLEUCINE	1		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCA	YES	PROLINE	1		
TCA	YES	SERINE	1	POLAR NONIONIZABLE (POL)	3
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	1		
		TYROSINE	0		
ACA	YES	THREONINE	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
		ASPARTIC ACID	0		
GAA	YES	GLUTAMIC ACID	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	1
AAA	YES	LYSINE	1		
		ARGININE	0		
		HISTIDINE	0	STOP SIGNAL (STP)	1
TAA	YES	STOP CODON	1		
TOTAL	12	10 Amino Acids Are Represented		NPL: 6: POL: 3: NEG: 1: POS: 1: STP: 1	

TABLE 41. Mutagenic Cassette: N, A/C/G, A

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGA	YES	GLYCINE	1	NONPOLAR (NPL)	3
GCA	YES	ALANINE	1		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCA	YES	PROLINE	1		
TCA	YES	SERINE	1	POLAR NONIONIZABLE (POL)	3
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	1		
		TYROSINE	0		
ACA	YES	THREONINE	1	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	1
		ASPARTIC ACID	0		
GAA	YES	GLUTAMIC ACID	1	IONIZABLE: BASIC POSITIVE CHARGE (POS)	3
AAA	YES	LYSINE	1		
CGA	YES	ARGININE	2		
AGA	YES			STOP SIGNAL (STP)	2
		HISTIDINE	0		
TAA	YES	STOP CODON	2	TOTAL	
TGA	YES				
	12	9 Amino Acids Are Represented		NPL: 3: POL: 3: NEG: 1: POS: 3: STP: 2	

TABLE 42. Mutagenic Cassette: A, N, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES				
		METHIONINE	1		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	8
AGC	YES				
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
AGA	YES	ARGININE	2		
AGG	YES				
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
	16	7 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 4: 8: 0: 4: 0	

TABLE 43. Mutagenic Cassette: C, N, N

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	8
			ALANINE	0		
			VALINE	0		
CTT	YES	LEUCINE		4		
CTC	YES					
CTA	YES					
CTG	YES					
		ISOLEUCINE	0			
		METHIONINE	0			
		PHENYLALANINE	0			
		TRYPTOPHAN	0			
CCT	YES	PROLINE		4		
CCC	YES					
CCA	YES					
CCG	YES					
		SERINE	0	POLAR NONIONIZABLE (POL)	2	
		CYSTEINE	0			
		ASPARAGINE	0			
CAA	YES	GLUTAMINE	2			
CAG	YES					
		TYROSINE	0			
		THREONINE	0			
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0	
		GLUTAMIC ACID	0			
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6	
CGT	YES	ARGININE	4			
CGC	YES					
CGA	YES					
CGG	YES					
CAT	YES	HISTIDINE	2			
CAC	YES					
		STOP CODON	0	STOP SIGNAL (STP)	0	
	16	5 Amino Acids Are Represented			NPL: POL: NEG: POS: STP = 8: 2: 0: 6: 0	

TABLE 44. Mutagenic Cassette: G, N, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	12
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2		
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
		LYSINE	0		
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0		

TABLE 45. Mutagenic Cassette: T, N, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	5
		ALANINE	0		
		VALINE	0		
TTA	YES	LEUCINE	2		
TTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	8
TCC	YES				
TCA	YES				
TCG	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
		ASPARAGINE	0		
		GLUTAMINE	0		
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
TOTAL	16	6 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 5: 8: 0: 0: 3	

TABLE 46. Mutagenic Cassette: A/C, N, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	12
		ALANINE	0		
		VALINE	0		
CTT	YES	LEUCINE	4		
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	10
AGC	YES				
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10
AAG	YES				
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
CAT	YES	HISTIDINE	2		
CAC	YES				
		STOP CODON	0	STOP SIGNAL (STP)	0
	32	11 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 12: 10: 0: 10: 0	

TABLE 47. Mutagenic Cassette: A/G, N, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	16
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
		LEUCINE	0		
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	8
AGC	YES				
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
AGA	YES	ARGININE	2		
AGG	YES				
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
	32	12 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 16: 8: 4: 4: 0	

TABLE 48. Mutagenic Cassette: A/T, N, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	9
		ALANINE	0		
		VALINE	0		
TTA	YES	LEUCINE	2		
TTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
TCT	YES	SERINE	6	POLAR NONIONIZABLE (POL)	16
TCC	YES				
TCA	YES				
TCG	YES				
AGT	YES				
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
		GLUTAMINE	0		
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
AGA	YES	ARGININE	2		
AGG	YES				
		HISTIDINE	0	STOP SIGNAL (STP)	3
TAA	YES	STOP CODON	3		
TAG	YES				
TGA	YES				
TOTAL	32	12 Amino Acids Are Represented		NPL: 9: POL: 16: NEG: 0: POS: 4: STP: 3	

TABLE 49. Mutagenic Cassette: C/G, N, N

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES		GLYCINE	4	NONPOLAR (NPL)	20
GGC	YES					
GGA	YES					
GGG	YES					
GCT	YES		ALANINE	4		
GCC	YES					
GCA	YES					
GCG	YES					
GTT	YES		VALINE	4		
GTC	YES					
GTA	YES					
GTG	YES					
CTT	YES		LEUCINE	4		
CTC	YES					
CTA	YES					
CTG	YES					
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
CCT	YES		PROLINE	4		
CCC	YES					
CCA	YES					
CCG	YES					
			SERINE	0		
			CYSTEINE	0		
			ASPARAGINE	0		
CAA	YES		GLUTAMINE	2		
CAG	YES					
			TYROSINE	0		
			THREONINE	0		
GAT	YES		ASPARTIC ACID	2		
GAC	YES					
GAA	YES		GLUTAMIC ACID	2		
GAG	YES					
			LYSINE	0		
CGT	YES		ARGININE	4		
CGC	YES					
CGA	YES					
CGG	YES					
CAT	YES		HISTIDINE	2		
CAC	YES					
			STOP CODON	0		
	32	10 Amino Acids Are Represented			NPL: POL: NEG: POS: STP =	
					20: 2: 4: 6: 0	

TABLE 50. Mutagenic Cassette: C/T, N, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	13
		ALANINE	0		
		VALINE	0		
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	10
TCC	YES				
TCA	YES				
TCG	YES				
TGT	YES	CYSTEINE	2		
TGC	YES	ASPARAGINE	0		
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES	THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT	YES	ARGININE	4		
CGC	YES				
CGA	YES				
CGG	YES				
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
TOTAL	32	10 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 13: 10: 0: 6: 3	

TABLE 51. Mutagenic Cassette: G/T, N, N

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT		YES	GLYCINE	4	NONPOLAR (NPL)	17
GGC		YES				
GGA		YES				
GGG		YES				
GCT		YES	ALANINE	4		
GCC		YES				
GCA		YES				
GCG		YES				
GTT		YES	VALINE	4		
GTC		YES				
GTA		YES				
GTG		YES				
TTA		YES	LEUCINE	2		
TTG		YES				
			ISOLEUCINE	0		
			METHIONINE	0		
TTT		YES	PHENYLALANINE	2		
TTC		YES				
TGG		YES	TRYPTOPHAN	1		
			PROLINE	0		
TCT		YES	SERINE	4	POLAR NONIONIZABLE (POL)	8
TCC		YES				
TCA		YES				
TCG		YES				
TGT		YES	CYSTEINE	2		
TGC		YES				
			ASPARAGINE	0		
			GLUTAMINE	0		
TAT		YES	TYROSINE	2		
TAC		YES				
			THREONINE	0		
GAT		YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC		YES				
GAA		YES	GLUTAMIC ACID	2		
GAG		YES				
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
			ARGININE	0		
			HISTIDINE	0		
TAA		YES	STOP CODON	3		
TAG		YES				
TGA		YES				
32		11 Amino Acids Are Represented			NPL: POL: NEG: POS: STP = 17: 8: 4: 0: 3	

TABLE 52. Mutagenic Cassette: N, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	6
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES				
TOTAL	16	7 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 0: 6: 4: 4: 2	

TABLE 53. Mutagenic Cassette: N, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4		
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	16	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 8: 8: 0: 0: 0	

TABLE 54. Mutagenic Cassette: N, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	5
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	4
AGC	YES				
TGT	YES				
TGC	YES				
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	16	5 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 5: 4: 0: 6: 1	

TABLE 55. Mutagenic Cassette: N, T, N

TABLE 55. Mutagenic Cassette 1, 2, 3					
CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	16
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0	POLAR NONIONIZABLE (POL)	0
		PROLINE	0		
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		GLUTAMIC ACID	0		
		LYSINE	0	STOP SIGNAL (STP)	0
		ARGININE	0		
		HISTIDINE	0	STOP SIGNAL (STP)	0
		STOP CODON	0		
TOTAL	16	5 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 16: 0: 0: 0: 0	

TABLE 56. Mutagenic Cassette: N, A/C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	14
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES				
TOTAL	32	11 Amino Acids Are Represented		NPL: 8: POL: 14: NEG: 4: POS: 4: STP = 2	

TABLE 57. Mutagenic Cassette: N, A/G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)		
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	5		
GGC	YES						
GGA	YES						
GGG	YES						
		ALANINE	0				
		VALINE	0				
		LEUCINE	0				
		ISOLEUCINE	0				
		METHIONINE	0				
		PHENYLALANINE	0				
TGG	YES	TRYPTOPHAN	1				
		PROLINE	0				
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	10		
AGC	YES						
TGT	YES	CYSTEINE	2				
TGC	YES						
AAT	YES	ASPARAGINE	2				
AAC	YES						
CAA	YES	GLUTAMINE	2				
CAG	YES						
TAT	YES	TYROSINE	2				
TAC	YES						
		THREONINE	0				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4		
GAC	YES						
GAA	YES	GLUTAMIC ACID	2				
GAG	YES						
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10		
AAG	YES						
CGT	YES	ARGININE	6				
CGC	YES						
CGA	YES						
CGG	YES						
AGA	YES						
AGG	YES						
CAT	YES	HISTIDINE	2				
CAC	YES						
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3		
TAG	YES						
TGA	YES						
TOTAL	32	12 Amino Acids Are Represented		NPL: 5: POL: 10: NEG: 4: POS: 10: STP = 3			

TABLE 58. Mutagenic Cassette: N, A/T, N

CODON	Represented	CATEGORY	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	16
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2	POLAR NONIONIZABLE (POL)	6
TTC	YES				
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0		
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
TAC	YES				
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2		
GAC	YES			IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2		
AAG	YES			STOP SIGNAL (STP)	2
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES			STOP SIGNAL (STP)	2
TAA	YES	STOP CODON	2		
TAG	YES			NPL: POL: NEG: POS: STP =	16: 6: 4: 4: 2
TOTAL	32	12 Amino Acids Are Represented			

TABLE 59. Mutagenic Cassette: N, C/G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	13
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	6	POLAR NONIONIZABLE (POL)	12
TCC	YES				
TCA	YES				
TCG	YES				
AGT	YES	CYSTEINE	2		
AGC	YES				
TGT	YES				
TGC	YES				
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
	32	8 Amino Acids Are Represented		NPL: 13: POL: 12: NEG: POS: STP = 0: 6: 1	

TABLE 60. Mutagenic Cassette: N, C/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	24
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	8
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
	32	9 Amino Acids Are Represented		NPL: 24: POL: 8: NEG: 0: POS: 0: STP = 0	

TABLE 61. Mutagenic Cassette: N, G/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	21
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	4
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
	32	10 Amino Acids Are Represented		NPL: 21: POL: 4: NEG: 0: POS: 6: STP = 1	

TABLE 62. Mutagenic Cassette: N, A/C/G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	13
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	6	POLAR NONIONIZABLE (POL)	18
TCC	YES				
TCA	YES				
TCG	YES				
AGT	YES	CYSTEINE	2		
AGC	YES				
TGT	YES				
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10
AAG	YES				
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES	HISTIDINE	2		
CAT	YES				
CAC	YES				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
	48	15 Amino Acids Are Represented		NPL: 13: POL: 18: NEG: 4: POS: 10: STP: 3	

TABLE 63. Mutagenic Cassette: N, A/C/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	24
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	14
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
GAG	YES				
AAA	YES	LYSINE	2		
AAG	YES				
		ARGININE	0	STOP SIGNAL (STP)	2
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	2		
TAG	YES				
TOTAL	48	16 Amino Acids Are Represented		NPL: 24: POL: 14: NEG:POS: 4: STP: 2	

TABLE 64. Mutagenic Cassette: N, A/G/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)		
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	21		
GGC	YES						
GGA	YES						
GGG	YES						
		ALANINE	0				
GTT	YES	VALINE	4				
GTC	YES						
GTA	YES						
GTG	YES						
TTA	YES	LEUCINE	6				
TTG	YES						
CTT	YES						
CTC	YES						
CTA	YES						
CTG	YES						
ATT	YES	ISOLEUCINE	3				
ATC	YES						
ATA	YES						
ATG	YES	METHIONINE	1				
TTT	YES	PHENYLALANINE	2				
TTC	YES						
TGG	YES	TRYPTOPHAN	1				
		PROLINE	0				
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	10		
AGC	YES						
TGT	YES	CYSTEINE	2				
TGC	YES						
AAT	YES	ASPARAGINE	2				
AAC	YES						
CAA	YES	GLUTAMINE	2				
CAG	YES						
TAT	YES	TYROSINE	2				
TAC	YES						
		THREONINE	0				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4		
GAC	YES						
GAA	YES	GLUTAMIC ACID	2				
GAG	YES						
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10		
AAG	YES						
CGT	YES	ARGININE	6				
CGC	YES						
CGA	YES						
CGG	YES						
AGA	YES						
AGG	YES						
CAT	YES	HISTIDINE	2				
CAC	YES						
TAA	YES	STOP CODON	3			STOP SIGNAL (STP)	3
TAG	YES						
TGA	YES						
TOTAL	48	17 Amino Acids Are Represented		NPL: 21: POL: 10: NEG:POS: 4: STP = 10: 3			

TABLE 65. Mutagenic Cassette: N, C/G/T, N

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT		YES	GLYCINE	4	NONPOLAR (NPL)	29
GGC		YES				
GGA		YES				
GGG		YES				
GCT		YES	ALANINE	4		
GCC		YES				
GCA		YES				
GCG		YES				
GTT		YES	VALINE	4		
GTC		YES				
GTA		YES				
GTG		YES				
TTA		YES	LEUCINE	6		
TTG		YES				
CTT		YES				
CTC		YES				
CTA		YES				
CTG		YES				
ATT		YES	ISOLEUCINE	3		
ATC		YES				
ATA		YES				
ATG		YES	METHIONINE	1		
TTT		YES	PHENYLALANINE	2		
TTC		YES				
TGG		YES	TRYPTOPHAN	1		
CCT		YES	PROLINE	4		
CCC		YES				
CCA		YES				
CCG		YES				
TCT		YES	SERINE	6	POLAR NONIONIZABLE (POL)	12
TCC		YES				
TCA		YES				
TCG		YES				
AGT		YES				
AGC		YES				
TGT		YES	CYSTEINE	2		
TGC		YES				
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
ACT		YES	THREONINE	4		
ACC		YES				
ACA		YES				
ACG		YES				
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT		YES	ARGININE	6		
CGC		YES				
CGA		YES				
CGG		YES				
AGA		YES				
AGG		YES				
			HISTIDINE	0		
TGA		YES	STOP CODON	1	STOP SIGNAL (STP)	1
		48	13 Amino Acids Are Represented		NPL: 29: POL: 12: NEG:POS: 0: STP: 1	

TABLE 66. Mutagenic Cassette: C, C, N

Mutagenic Cassette C, C, T, T					
CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	4		
CCT	YES			POLAR NONIONIZABLE (POL)	0
CCC	YES				
CCA	YES				
CCG	YES				
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ASPARTIC ACID	0		
		GLUTAMIC ACID	0		
		LYSINE	0	STOP SIGNAL (STP)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0		
	4	1 Amino Acid Is Represented		NPL: POL: NEG:POS: STP = 4: 0: 0: 0: 0	

TABLE 67. Mutagenic Cassette: G, G, N

Initiation Cassette: G, G, N					
CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	4
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
0	4	1 Amino Acid Is Represented		NPL: POL: NEG:POS: STP = 4: 0: 0: 0: 0	

TABLE 68. Mutagenic Cassette: G, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0	POLAR NONIONIZABLE (POL)	0
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		ASPARTIC ACID	0		
		GLUTAMIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		LYSINE	0		
		ARGININE	0		
		HISTIDINE	0	STOP SIGNAL (STP)	0
		STOP CODON	0		
TOTAL	4	1 Amino Acid Is Represented		NPL: 4: POL: 0: NEG:POS: STP - 0: 0: 0	

TABLE 69. Mutagenic Cassette: G, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0	POLAR NONIONIZABLE (POL)	0
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: 4: POL: 0: NEG: 0: POS: 0: STP: 0:	

TABLE 70. Mutagenic Cassette: C, G, N

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	0
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
			SERINE	0	POLAR NONIONIZABLE (POL)	0
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
CGT	YES	ARGININE		4		
CGC	YES					
CGA	YES					
CGG	YES					
			HISTIDINE	0	STOP SIGNAL (STP)	0
			STOP CODON	0		
TOTAL		4	1 Amino Acid Is Represented		NPL: 0: POL: 0: NEG:POS: 0: STP: 4: 0	

TABLE 71. Mutagenic Cassette: C, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
CTT	YES	LEUCINE	4		
CTC	YES				
CTA	YES				
CTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0	POLAR NONIONIZABLE (POL)	0
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: POL: NEG:POS: STP = 4: 0: 0: 0: 0	

TABLE 72. Mutagenic Cassette: T, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	4
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: 0: POL: 4: NEG:POS: STP = 0: 0: 0: 0	

TABLE 73. Mutagenic Cassette: A, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	4
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	4		
ACT	YES			IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
ACC	YES				
ACA	YES			IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
ACG	YES				
		ASPARTIC ACID	0		
		GLUTAMIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		LYSINE	0		
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: 0: POL: 4: NEG:POS: STP = 0: 0: 0: 0	

TABLE 74. Mutagenic Cassette: G, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0		
		CYSTEINE	0	POLAR NONIONIZABLE (POL)	0
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
GAG	YES				
		LYSINE	0		
		ARGININE	0	STOP SIGNAL (STP)	0
		HISTIDINE	0		
		STOP CODON	0		
TOTAL	4	2 Amino Acids Are Represented		NPL: 0: POL: 0: NEG:POS: STP = 4: 0: 0	

TABLE 75. Mutagenic Cassette: A, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0	POLAR NONIONIZABLE (POL)	0
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		GLUTAMIC ACID	0		
		LYSINE	0		
		ARGININE	0	STOP SIGNAL (STP)	0
		HISTIDINE	0		
		STOP CODON	0		
TOTAL	4	2 Amino Acids Are Represented		NPL: 4: POL: 0: NEG:POS: STP = 0: 0: 0	

TABLE 76. Mutagenic Cassette: C, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0		
		CYSTEINE	0	POLAR NONIONIZABLE (POL)	2
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	2		
CAG	YES	TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: 0: POL: 2: NEG: 0: POS: 2: STP: 0	

TABLE 77. Mutagenic Cassette: T, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
TTA	YES	LEUCINE	2		
TTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
		PROLINE	0	POLAR NONIONIZABLE (POL)	0
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: 4: POL: 0: NEG: 0: POS: 0: STP: 0	

TABLE 78. Mutagenic Cassette: A, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		ASPARTIC ACID	0		
		GLUTAMIC ACID	0		
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
AAG	YES				
		ARGININE	0		
		HISTIDINE	0	STOP SIGNAL (STP)	0
		STOP CODON	0		
TOTAL	4	2 Amino Acids Are Represented		NPL: 0: POL: 2: NEG:POS: STP = 0: 2: 0: 0	

TABLE 79. Mutagenic Cassette: T, A, N

CODON		Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	0
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
			SERINE	0	POLAR NONIONIZABLE (POL)	2
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
TAT	YES		TYROSINE	2		
TAC	YES					
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
			ARGININE	0		
			HISTIDINE	0		
TAA	YES		STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES					
TOTAL		4	1 Amino Acid Is Represented		NPL: 0: POL: 2: NEG:POS: STP = 0: 2: 0: 0: 2	

TABLE 80. Mutagenic Cassette: T, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	1
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0	POLAR NONIONIZABLE (POL)	2
		SERINE	0		
TGT	YES	CYSTEINE	2		
TGC	YES				
		TYROSINE	0		
		THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		ASPARTIC ACID	0		
		GLUTAMIC ACID	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		LYSINE	0		
		ARGININE	0		
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	4	2 Amino Acids Are Represented		NPL: 1 POL: 2 NEG: 0 POS: 0 STP = 1	

TABLE 81. Mutagenic Cassette: A, G, N

Multiple Choice: A, G, N					
CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0	POLAR NONIONIZABLE (POL)	2
AGT	YES	SERINE	2		
AGC	YES				
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
AGA	YES	ARGININE	2		
AGG	YES				
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: 0 POL: 2 NEG: 0 POS: 2 STP = 0	

TABLE 82. Mutagenic Cassette: G/C, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	4
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0		
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0		
		GLUTAMIC ACID	0		
		LYSINE	0		
		ARGININE	4	POLAR NONIONIZABLE (POL)	0
CGT	YES				
CGC	YES				
CGA	YES				
CGG	YES				
		HISTIDINE	0		
		STOP CODON	0		
				IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
				IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
				STOP SIGNAL (STP)	0
TOTAL	8	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 4: 0: 0: 4: 0	

TABLE 83. Mutagenic Cassette: G/C, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		ASPARTIC ACID	0		
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	8	2 Amino Acids Are Represented		NPL: POL: NEG: POS:STP = 8: 0: 0: 0: 0	

TABLE 84. Mutagenic Cassette: G/C, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	2		
CAG	YES				
		TYROSINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2		
GAC	YES				
GAA	YES	GLUTAMIC ACID	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
GAG	YES				
		LYSINE	0		
		ARGININE	0		
CAT	YES	HISTIDINE	2	STOP SIGNAL (STP)	0
CAC	YES				
		STOP CODON	0		
TOTAL	8	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 0: 2: 4: 2: 0	

TABLE 85. Mutagenic Cassette: G/C, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)		
		GLYCINE	0	NONPOLAR (NPL)	8		
		ALANINE	0				
GTT	YES	VALINE	4				
GTC	YES						
GTA	YES						
GTG	YES						
CTT	YES	LEUCINE	4				
CTC	YES						
CTA	YES						
CTG	YES						
		ISOLEUCINE	0	POLAR NONIONIZABLE (POL)	0		
		METHIONINE	0				
		PHENYLALANINE	0				
		TRYPTOPHAN	0				
		PROLINE	0				
		SERINE	0				
		CYSTEINE	0				
		ASPARAGINE	0				
		GLUTAMINE	0				
		TYROSINE	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0		
		THREONINE	0				
		ASPARTIC ACID	0				
		GLUTAMIC ACID	0				
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0		
		ARGININE	0				
		HISTIDINE	0				
		STOP CODON	0	STOP SIGNAL (STP)	0		
TOTAL	8	2 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 8: 0: 0: 0: 0			

2.11.2. CHIMERIZATIONS

2.11.2.1 "SHUFFLING"

5 Nucleic acid shuffling is a method for *in vitro* or *in vivo* homologous recombination of pools of shorter or smaller polynucleotides to produce a polynucleotide or polynucleotides. Mixtures of related nucleic acid sequences or polynucleotides are subjected to sexual PCR to provide random polynucleotides, and reassembled to yield a library or mixed population of recombinant hybrid nucleic acid molecules or
10 polynucleotides.

 In contrast to cassette mutagenesis, only shuffling and error-prone PCR allow one to mutate a pool of sequences blindly (without sequence information other than primers).

15 The advantage of the mutagenic shuffling of this invention over error-prone PCR alone for repeated selection can best be explained with an example from antibody engineering. Consider DNA shuffling as compared with error-prone PCR (not sexual PCR). The initial library of selected pooled sequences can consist of related sequences of diverse origin (i.e. antibodies from naive mRNA) or can be derived by any type of
20 mutagenesis (including shuffling) of a single antibody gene. A collection of selected complementarity determining regions ("CDRs") is obtained after the first round of affinity selection. In the diagram the thick CDRs confer onto the antibody molecule increased affinity for the antigen. Shuffling allows the free combinatorial association of all of the CDR1s with all of the CDR2s with all of the CDR3s, for example.

25

 This method differs from error-prone PCR, in that it is an inverse chain reaction. In error-prone PCR, the number of polymerase start sites and the number of molecules grows exponentially. However, the sequence of the polymerase start sites and the sequence of the molecules remains essentially the same. In contrast, in nucleic acid
30 reassembly or shuffling of random polynucleotides the number of start sites and the

number (but not size) of the random polynucleotides decreases over time. For polynucleotides derived from whole plasmids the theoretical endpoint is a single, large concatemeric molecule.

5 Since cross-overs occur at regions of homology, recombination will primarily occur between members of the same sequence family. This discourages combinations of CDRs that are grossly incompatible (e.g., directed against different epitopes of the same antigen). It is contemplated that multiple families of sequences can be shuffled in the same reaction. Further, shuffling generally conserves the relative order, such that, for
10 example, CDR1 will not be found in the position of CDR2.

Rare shufflants will contain a large number of the best (eg. highest affinity) CDRs and these rare shufflants may be selected based on their superior affinity.

15 CDRs from a pool of 100 different selected antibody sequences can be permuted in up to 1006 different ways. This large number of permutations cannot be represented in a single library of DNA sequences. Accordingly, it is contemplated that multiple cycles of DNA shuffling and selection may be required depending on the length of the sequence and the sequence diversity desired.

20

Error-prone PCR, in contrast, keeps all the selected CDRs in the same relative sequence, generating a much smaller mutant cloud.

25 The template polynucleotide which may be used in the methods of this invention may be DNA or RNA. It may be of various lengths depending on the size of the gene or shorter or smaller polynucleotide to be recombined or reassembled. Preferably, the template polynucleotide is from 50 bp to 50 kb. It is contemplated that entire vectors containing the nucleic acid encoding the protein of interest can be used in the methods of this invention, and in fact have been successfully used.

The template polynucleotide may be obtained by amplification using the PCR reaction (USPN 4,683,202 and USPN 4,683,195) or other amplification or cloning methods. However, the removal of free primers from the PCR products before subjecting
5 them to pooling of the PCR products and sexual PCR may provide more efficient results. Failure to adequately remove the primers from the original pool before sexual PCR can lead to a low frequency of crossover clones.

The template polynucleotide often should be double-stranded. A double-stranded
10 nucleic acid molecule is recommended to ensure that regions of the resulting single-stranded polynucleotides are complementary to each other and thus can hybridize to form a double-stranded molecule.

It is contemplated that single-stranded or double-stranded nucleic acid
15 polynucleotides having regions of identity to the template polynucleotide and regions of heterology to the template polynucleotide may be added to the template polynucleotide, at this step. It is also contemplated that two different but related polynucleotide templates can be mixed at this step.

20 The double-stranded polynucleotide template and any added double-or single-stranded polynucleotides are subjected to sexual PCR which includes slowing or halting to provide a mixture of from about 5 bp to 5 kb or more. Preferably the size of the random polynucleotides is from about 10 bp to 1000 bp, more preferably the size of the polynucleotides is from about 20 bp to 500 bp.

25

Alternatively, it is also contemplated that double-stranded nucleic acid having multiple nicks may be used in the methods of this invention. A nick is a break in one strand of the double-stranded nucleic acid. The distance between such nicks is preferably 5 bp to 5 kb, more preferably between 10 bp to 1000 bp. This can provide areas of self-

priming to produce shorter or smaller polynucleotides to be included with the polynucleotides resulting from random primers, for example.

5 The concentration of any one specific polynucleotide will not be greater than 1% by weight of the total polynucleotides, more preferably the concentration of any one specific nucleic acid sequence will not be greater than 0.1% by weight of the total nucleic acid.

10 The number of different specific polynucleotides in the mixture will be at least about 100, preferably at least about 500, and more preferably at least about 1000.

At this step single-stranded or double-stranded polynucleotides, either synthetic or natural, may be added to the random double-stranded shorter or smaller polynucleotides in order to increase the heterogeneity of the mixture of polynucleotides.

15

It is also contemplated that populations of double-stranded randomly broken polynucleotides may be mixed or combined at this step with the polynucleotides from the sexual PCR process and optionally subjected to one or more additional sexual PCR cycles.

20

Where insertion of mutations into the template polynucleotide is desired, single-stranded or double-stranded polynucleotides having a region of identity to the template polynucleotide and a region of heterology to the template polynucleotide may be added in a 20 fold excess by weight as compared to the total nucleic acid, more preferably the single-stranded polynucleotides may be added in a 10 fold excess by weight as compared to the total nucleic acid.

25

Where a mixture of different but related template polynucleotides is desired, populations of polynucleotides from each of the templates may be combined at a ratio of

less than about 1:100, more preferably the ratio is less than about 1:40. For example, a backcross of the wild-type polynucleotide with a population of mutated polynucleotide may be desired to eliminate neutral mutations (e.g., mutations yielding an insubstantial alteration in the phenotypic property being selected for). In such an example, the ratio of
5 randomly provided wild-type polynucleotides which may be added to the randomly provided sexual PCR cycle hybrid polynucleotides is approximately 1:1 to about 100:1, and more preferably from 1:1 to 40:1.

The mixed population of random polynucleotides are denatured to form
10 single-stranded polynucleotides and then re-annealed. Only those single-stranded polynucleotides having regions of homology with other single-stranded polynucleotides will re-anneal.

The random polynucleotides may be denatured by heating. One skilled in the art
15 could determine the conditions necessary to completely denature the double-stranded nucleic acid. Preferably the temperature is from 80 °C to 100 °C, more preferably the temperature is from 90 °C to 96 °C. other methods which may be used to denature the polynucleotides include pressure (36) and pH.

20 The polynucleotides may be re-annealed by cooling. Preferably the temperature is from 20 °C to 75 °C, more preferably the temperature is from 40 °C to 65 °C. If a high frequency of crossovers is needed based on an average of only 4 consecutive bases of homology, recombination can be forced by using a low annealing temperature, although the process becomes more difficult. The degree of renaturation which occurs will depend
25 on the degree of homology between the population of single-stranded polynucleotides.

Renaturation can be accelerated by the addition of polyethylene glycol ("PEG") or salt. The salt concentration is preferably from 0 mM to 200 mM, more preferably the salt

concentration is from 10 mM to 100 mM. The salt may be KCl or NaCl. The concentration of PEG is preferably from 0% to 20%, more preferably from 5% to 10%.

5 The annealed polynucleotides are next incubated in the presence of a nucleic acid polymerase and dNTP's (i.e. dATP, dCTP, dGTP and dTTP). The nucleic acid polymerase may be the Klenow fragment, the Taq polymerase or any other DNA polymerase known in the art.

10 The approach to be used for the assembly depends on the minimum degree of homology that should still yield crossovers. If the areas of identity are large, Taq polymerase can be used with an annealing temperature of between 45-65 °C. If the areas of identity are small, Klenow polymerase can be used with an annealing temperature of between 20-30 °C. One skilled in the art could vary the temperature of annealing to increase the number of cross-overs achieved.

15 The polymerase may be added to the random polynucleotides prior to annealing, simultaneously with annealing or after annealing.

20 The cycle of denaturation, renaturation and incubation in the presence of polymerase is referred to herein as shuffling or reassembly of the nucleic acid. This cycle is repeated for a desired number of times. Preferably the cycle is repeated from 2 to 50 times, more preferably the sequence is repeated from 10 to 40 times.

25 The resulting nucleic acid is a larger double-stranded polynucleotide of from about 50 bp to about 100 kb, preferably the larger polynucleotide is from 500 bp to 50 kb.

This larger polynucleotides may contain a number of copies of a polynucleotide having the same size as the template polynucleotide in tandem. This concatemeric polynucleotide is then denatured into single copies of the template polynucleotide. The

result will be a population of polynucleotides of approximately the same size as the template polynucleotide. The population will be a mixed population where single or double-stranded polynucleotides having an area of identity and an area of heterology have been added to the template polynucleotide prior to shuffling. These polynucleotides
5 are then cloned into the appropriate vector and the ligation mixture used to transform bacteria.

It is contemplated that the single polynucleotides may be obtained from the larger concatemeric polynucleotide by amplification of the single polynucleotide prior to
10 cloning by a variety of methods including PCR (USPN 4,683,195 and USPN 4,683,202), rather than by digestion of the concatemer.

The vector used for cloning is not critical provided that it will accept a polynucleotide of the desired size. If expression of the particular polynucleotide is
15 desired, the cloning vehicle should further comprise transcription and translation signals next to the site of insertion of the polynucleotide to allow expression of the polynucleotide in the host cell. Preferred vectors include the pUC series and the pBR series of plasmids.

20 The resulting bacterial population will include a number of recombinant polynucleotides having random mutations. This mixed population may be tested to identify the desired recombinant polynucleotides. The method of selection will depend on the polynucleotide desired.

25 For example, if a polynucleotide which encodes a protein with increased binding efficiency to a ligand is desired, the proteins expressed by each of the portions of the polynucleotides in the population or library may be tested for their ability to bind to the ligand by methods known in the art (i.e. panning, affinity chromatography). If a polynucleotide which encodes for a protein with increased drug resistance is desired, the

proteins expressed by each of the polynucleotides in the population or library may be tested for their ability to confer drug resistance to the host organism. One skilled in the art, given knowledge of the desired protein, could readily test the population to identify polynucleotides which confer the desired properties onto the protein.

5

It is contemplated that one skilled in the art could use a phage display system in which fragments of the protein are expressed as fusion proteins on the phage surface (Pharmacia, Milwaukee WI). The recombinant DNA molecules are cloned into the phage DNA at a site which results in the transcription of a fusion protein a portion of which is encoded by the recombinant DNA molecule. The phage containing the recombinant nucleic acid molecule undergoes replication and transcription in the cell. The leader sequence of the fusion protein directs the transport of the fusion protein to the tip of the phage particle. Thus the fusion protein which is partially encoded by the recombinant DNA molecule is displayed on the phage particle for detection and selection by the methods described above.

10
15

It is further contemplated that a number of cycles of nucleic acid shuffling may be conducted with polynucleotides from a sub-population of the first population, which sub-population contains DNA encoding the desired recombinant protein. In this manner, proteins with even higher binding affinities or enzymatic activity could be achieved.

20

It is also contemplated that a number of cycles of nucleic acid shuffling may be conducted with a mixture of wild-type polynucleotides and a sub-population of nucleic acid from the first or subsequent rounds of nucleic acid shuffling in order to remove any silent mutations from the sub-population.

25

Any source of nucleic acid, in purified form can be utilized as the starting nucleic acid. Thus the process may employ DNA or RNA including messenger RNA, which DNA or RNA may be single or double stranded. In addition, a DNA-RNA hybrid which

contains one strand of each may be utilized. The nucleic acid sequence may be of various lengths depending on the size of the nucleic acid sequence to be mutated. Preferably the specific nucleic acid sequence is from 50 to 50000 base pairs. It is contemplated that entire vectors containing the nucleic acid encoding the protein of
5 interest may be used in the methods of this invention.

The nucleic acid may be obtained from any source, for example, from plasmids such a pBR322, from cloned DNA or RNA or from natural DNA or RNA from any source including bacteria, yeast, viruses and higher organisms such as plants or animals.
10 DNA or RNA may be extracted from blood or tissue material. The template polynucleotide may be obtained by amplification using the polynucleotide chain reaction (PCR, see USPN 4,683,202 and USPN 4,683,195). Alternatively, the polynucleotide may be present in a vector present in a cell and sufficient nucleic acid may be obtained by culturing the cell and extracting the nucleic acid from the cell by methods known in the
15 art.

Any specific nucleic acid sequence can be used to produce the population of hybrids by the present process. It is only necessary that a small population of hybrid sequences of the specific nucleic acid sequence exist or be created prior to the present
20 process.

The initial small population of the specific nucleic acid sequences having mutations may be created by a number of different methods. Mutations may be created by error-prone PCR. Error-prone PCR uses low-fidelity polymerization conditions to
25 introduce a low level of point mutations randomly over a long sequence. Alternatively, mutations can be introduced into the template polynucleotide by oligonucleotide-directed mutagenesis. In oligonucleotide-directed mutagenesis, a short sequence of the polynucleotide is removed from the polynucleotide using restriction enzyme digestion and is replaced with a synthetic polynucleotide in which various bases have been altered

from the original sequence. The polynucleotide sequence can also be altered by chemical mutagenesis. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, hydroxylamine, hydrazine or formic acid. other agents which are analogues of nucleotide precursors include nitrosoguanidine, 5-bromouracil, 2-aminopurine, or acridine.

- 5 Generally, these agents are added to the PCR reaction in place of the nucleotide precursor thereby mutating the sequence. Intercalating agents such as proflavine, acriflavine, quinacrine and the like can also be used. Random mutagenesis of the polynucleotide sequence can also be achieved by irradiation with X-rays or ultraviolet light. Generally, plasmid polynucleotides so mutagenized are introduced into *E. coli* and propagated as a
- 10 pool or library of hybrid plasmids.

- Alternatively the small mixed population of specific nucleic acids may be found in nature in that they may consist of different alleles of the same gene or the same gene from different related species (i.e., cognate genes). Alternatively, they may be related
- 15 DNA sequences found within one species, for example, the immunoglobulin genes.

- Once the mixed population of the specific nucleic acid sequences is generated, the polynucleotides can be used directly or inserted into an appropriate cloning vector, using techniques well-known in the art.

20

- The choice of vector depends on the size of the polynucleotide sequence and the host cell to be employed in the methods of this invention. The templates of this invention may be plasmids, phages, cosmids, phagemids, viruses (e.g., retroviruses, parainfluenzavirus, herpesviruses, reoviruses, paramyxoviruses, and the like), or selected
- 25 portions thereof (e.g., coat protein, spike glycoprotein, capsid protein). For example, cosmids and phagemids are preferred where the specific nucleic acid sequence to be mutated is larger because these vectors are able to stably propagate large polynucleotides.

If the mixed population of the specific nucleic acid sequence is cloned into a vector it can be clonally amplified by inserting each vector into a host cell and allowing the host cell to amplify the vector. This is referred to as clonal amplification because while the absolute number of nucleic acid sequences increases, the number of hybrids
5 does not increase. Utility can be readily determined by screening expressed polypeptides.

The DNA shuffling method of this invention can be performed blindly on a pool of unknown sequences. By adding to the reassembly mixture oligonucleotides (with ends that are homologous to the sequences being reassembled) any sequence mixture can be
10 incorporated at any specific position into another sequence mixture. Thus, it is contemplated that mixtures of synthetic oligonucleotides, PCR polynucleotides or even whole genes can be mixed into another sequence library at defined positions. The insertion of one sequence (mixture) is independent from the insertion of a sequence in another part of the template. Thus, the degree of recombination, the homology required,
15 and the diversity of the library can be independently and simultaneously varied along the length of the reassembled DNA.

This approach of mixing two genes may be useful for the humanization of antibodies from murine hybridomas. The approach of mixing two genes or inserting
20 alternative sequences into genes may be useful for any therapeutically used protein, for example, interleukin I, antibodies, tPA and growth hormone. The approach may also be useful in any nucleic acid for example, promoters or introns or 3' untranslated region or 5' untranslated regions of genes to increase expression or alter specificity of expression of proteins. The approach may also be used to mutate ribozymes or aptamers.

25

Shuffling requires the presence of homologous regions separating regions of diversity. Scaffold-like protein structures may be particularly suitable for shuffling. The conserved scaffold determines the overall folding by self-association, while displaying relatively unrestricted loops that mediate the specific binding. Examples of such

scaffolds are the immunoglobulin beta-barrel, and the four-helix bundle which are well-known in the art. This shuffling can be used to create scaffold-like proteins with various combinations of mutated sequences for binding.

5

In vitro Shuffling

The equivalents of some standard genetic matings may also be performed by shuffling *in vitro*. For example, a "molecular backcross" can be performed by repeatedly
10 mixing the hybrid's nucleic acid with the wild-type nucleic acid while selecting for the mutations of interest. As in traditional breeding, this approach can be used to combine phenotypes from different sources into a background of choice. It is useful, for example, for the removal of neutral mutations that affect unselected characteristics (i.e. immunogenicity). Thus it can be useful to determine which mutations in a protein are
15 involved in the enhanced biological activity and which are not, an advantage which cannot be achieved by error-prone mutagenesis or cassette mutagenesis methods.

Large, functional genes can be assembled correctly from a mixture of small random polynucleotides. This reaction may be of use for the reassembly of genes from
20 the highly fragmented DNA of fossils. In addition random nucleic acid fragments from fossils may be combined with polynucleotides from similar genes from related species.

It is also contemplated that the method of this invention can be used for the *in vitro* amplification of a whole genome from a single cell as is needed for a variety of
25 research and diagnostic applications. DNA amplification by PCR is in practice limited to a length of about 40 kb. Amplification of a whole genome such as that of *E. coli* (5, 000 kb) by PCR would require about 250 primers yielding 125 forty kb polynucleotides. This approach is not practical due to the unavailability of sufficient sequence data. On the other hand, random production of polynucleotides of the genome with sexual PCR cycles,
30 followed by gel purification of small polynucleotides will provide a multitude of possible

primers. Use of this mix of random small polynucleotides as primers in a PCR reaction alone or with the whole genome as the template should result in an inverse chain reaction with the theoretical endpoint of a single concatamer containing many copies of the genome.

5

100 fold amplification in the copy number and an average polynucleotide size of greater than 50 kb may be obtained when only random polynucleotides are used. It is thought that the larger concatamer is generated by overlap of many smaller polynucleotides. The quality of specific PCR products obtained using synthetic primers will be indistinguishable from the product obtained from unamplified DNA. It is expected that this approach will be useful for the mapping of genomes.

The polynucleotide to be shuffled can be produced as random or non-random polynucleotides, at the discretion of the practitioner. Moreover, this invention provides a method of shuffling that is applicable to a wide range of polynucleotide sizes and types, including the step of generating polynucleotide monomers to be used as building blocks in the reassembly of a larger polynucleotide. For example, the building blocks can be fragments of genes or they can be comprised of entire genes or gene pathways, or any combination thereof.

20

In vivo Shuffling

In an embodiment of *in vivo* shuffling, the mixed population of the specific nucleic acid sequence is introduced into bacterial or eukaryotic cells under conditions such that at least two different nucleic acid sequences are present in each host cell. The polynucleotides can be introduced into the host cells by a variety of different methods. The host cells can be transformed with the smaller polynucleotides using methods known in the art, for example treatment with calcium chloride. If the polynucleotides are inserted into a phage genome, the host cell can be transfected with the recombinant phage

25

genome having the specific nucleic acid sequences. Alternatively, the nucleic acid sequences can be introduced into the host cell using electroporation, transfection, lipofection, biolistics, conjugation, and the like.

- 5 In general, in this embodiment, the specific nucleic acids sequences will be present in vectors which are capable of stably replicating the sequence in the host cell. In addition, it is contemplated that the vectors will encode a marker gene such that host cells having the vector can be selected. This ensures that the mutated specific nucleic acid sequence can be recovered after introduction into the host cell. However, it is
- 10 contemplated that the entire mixed population of the specific nucleic acid sequences need not be present on a vector sequence. Rather only a sufficient number of sequences need be cloned into vectors to ensure that after introduction of the polynucleotides into the host cells each host cell contains one vector having at least one specific nucleic acid sequence present therein. It is also contemplated that rather than having a subset of the population
- 15 of the specific nucleic acids sequences cloned into vectors, this subset may be already stably integrated into the host cell.

- It has been found that when two polynucleotides which have regions of identity are inserted into the host cells homologous recombination occurs between the two
- 20 polynucleotides. Such recombination between the two mutated specific nucleic acid sequences will result in the production of double or triple hybrids in some situations.

- It has also been found that the frequency of recombination is increased if some of the mutated specific nucleic acid sequences are present on linear nucleic acid molecules.
- 25 Therefore, in a preferred embodiment, some of the specific nucleic acid sequences are present on linear polynucleotides.

 After transformation, the host cell transformants are placed under selection to identify those host cell transformants which contain mutated specific nucleic acid

sequences having the qualities desired. For example, if increased resistance to a particular drug is desired then the transformed host cells may be subjected to increased concentrations of the particular drug and those transformants producing mutated proteins able to confer increased drug resistance will be selected. If the enhanced ability of a particular protein to bind to a receptor is desired, then expression of the protein can be induced from the transformants and the resulting protein assayed in a ligand binding assay by methods known in the art to identify that subset of the mutated population which shows enhanced binding to the ligand. Alternatively, the protein can be expressed in another system to ensure proper processing.

10

Once a subset of the first recombined specific nucleic acid sequences (daughter sequences) having the desired characteristics are identified, they are then subject to a second round of recombination.

15

In the second cycle of recombination, the recombined specific nucleic acid sequences may be mixed with the original mutated specific nucleic acid sequences (parent sequences) and the cycle repeated as described above. In this way a set of second recombined specific nucleic acids sequences can be identified which have enhanced characteristics or encode for proteins having enhanced properties. This cycle can be repeated a number of times as desired.

20

It is also contemplated that in the second or subsequent recombination cycle, a backcross can be performed. A molecular backcross can be performed by mixing the desired specific nucleic acid sequences with a large number of the wild-type sequence, such that at least one wild-type nucleic acid sequence and a mutated nucleic acid sequence are present in the same host cell after transformation. Recombination with the wild-type specific nucleic acid sequence will eliminate those neutral mutations that may affect unselected characteristics such as immunogenicity but not the selected characteristics.

25

In another embodiment of this invention, it is contemplated that during the first round a subset of the specific nucleic acid sequences can be generated as smaller polynucleotides by slowing or halting their PCR amplification prior to introduction into the host cell. The size of the polynucleotides must be large enough to contain some regions of identity with the other sequences so as to homologously recombine with the other sequences. The size of the polynucleotides will range from 0.03 kb to 100 kb more preferably from 0.2 kb to 10 kb. It is also contemplated that in subsequent rounds, all of the specific nucleic acid sequences other than the sequences selected from the previous round may be utilized to generate PCR polynucleotides prior to introduction into the host cells.

The shorter polynucleotide sequences can be single-stranded or double-stranded. If the sequences were originally single-stranded and have become double-stranded they can be denatured with heat, chemicals or enzymes prior to insertion into the host cell. The reaction conditions suitable for separating the strands of nucleic acid are well known in the art.

The steps of this process can be repeated indefinitely, being limited only by the number of possible hybrids which can be achieved. After a certain number of cycles, all possible hybrids will have been achieved and further cycles are redundant.

In an embodiment the same mutated template nucleic acid is repeatedly recombined and the resulting recombinants selected for the desired characteristic.

Therefore, the initial pool or population of mutated template nucleic acid is cloned into a vector capable of replicating in a bacteria such as *E. coli*. The particular vector is not essential, so long as it is capable of autonomous replication in *E. coli*. In a preferred embodiment, the vector is designed to allow the expression and production of

any protein encoded by the mutated specific nucleic acid linked to the vector. It is also preferred that the vector contain a gene encoding for a selectable marker.

5 The population of vectors containing the pool of mutated nucleic acid sequences is introduced into the *E. coli* host cells. The vector nucleic acid sequences may be introduced by transformation, transfection or infection in the case of phage. The concentration of vectors used to transform the bacteria is such that a number of vectors is introduced into each cell. Once present in the cell, the efficiency of homologous recombination is such that homologous recombination occurs between the various
10 vectors. This results in the generation of hybrids (daughters) having a combination of mutations which differ from the original parent mutated sequences.

The host cells are then clonally replicated and selected for the marker gene present on the vector. Only those cells having a plasmid will grow under the selection.
15

The host cells which contain a vector are then tested for the presence of favorable mutations. Such testing may consist of placing the cells under selective pressure, for example, if the gene to be selected is an improved drug resistance gene. If the vector allows expression of the protein encoded by the mutated nucleic acid sequence, then such
20 selection may include allowing expression of the protein so encoded, isolation of the protein and testing of the protein to determine whether, for example, it binds with increased efficiency to the ligand of interest.

Once a particular daughter mutated nucleic acid sequence has been identified
25 which confers the desired characteristics, the nucleic acid is isolated either already linked to the vector or separated from the vector. This nucleic acid is then mixed with the first or parent population of nucleic acids and the cycle is repeated.

It has been shown that by this method nucleic acid sequences having enhanced desired properties can be selected.

In an alternate embodiment, the first generation of hybrids are retained in the cells and the parental mutated sequences are added again to the cells. Accordingly, the first
5 cycle of Embodiment I is conducted as described above. However, after the daughter nucleic acid sequences are identified, the host cells containing these sequences are retained.

The parent mutated specific nucleic acid population, either as polynucleotides or
10 cloned into the same vector is introduced into the host cells already containing the daughter nucleic acids. Recombination is allowed to occur in the cells and the next generation of recombinants, or granddaughters are selected by the methods described above.

15 This cycle can be repeated a number of times until the nucleic acid or peptide having the desired characteristics is obtained. It is contemplated that in subsequent cycles, the population of mutated sequences which are added to the preferred hybrids may come from the parental hybrids or any subsequent generation.

20 In an alternative embodiment, the invention provides a method of conducting a "molecular" backcross of the obtained recombinant specific nucleic acid in order to eliminate any neutral mutations. Neutral mutations are those mutations which do not confer onto the nucleic acid or peptide the desired properties. Such mutations may however confer on the nucleic acid or peptide undesirable characteristics. Accordingly, it
25 is desirable to eliminate such neutral mutations. The method of this invention provide a means of doing so.

In this embodiment, after the hybrid nucleic acid, having the desired characteristics, is obtained by the methods of the embodiments, the nucleic acid, the

vector having the nucleic acid or the host cell containing the vector and nucleic acid is isolated.

5 The nucleic acid or vector is then introduced into the host cell with a large excess of the wild-type nucleic acid. The nucleic acid of the hybrid and the nucleic acid of the wild-type sequence are allowed to recombine. The resulting recombinants are placed under the same selection as the hybrid nucleic acid. Only those recombinants which retained the desired characteristics will be selected. Any silent mutations which do not provide the desired characteristics will be lost through recombination with the wild-type
10 DNA. This cycle can be repeated a number of times until all of the silent mutations are eliminated.

Thus the methods of this invention can be used in a molecular backcross to eliminate unnecessary or silent mutations.

15

2.11.2.3. EXONUCLEASE-MEDIATED REASSEMBLY

In a particular embodiment, this invention provides for a method for shuffling,
20 assembling, reassembling, recombining, &/or concatenating at least two polynucleotides to form a progeny polynucleotide (e.g. a chimeric progeny polynucleotide that can be expressed to produce a polypeptide or a gene pathway). In a particular embodiment, a double stranded polynucleotide end (e.g. two single stranded sequences hybridized to each other as hybridization partners) is treated with an exonuclease to liberate nucleotides
25 from one of the two strands, leaving the remaining strand free of its original partner so that, if desired, the remaining strand may be used to achieve hybridization to another partner.

In a particular aspect, a double stranded polynucleotide end (that may be part of -
30 or connected to - a polynucleotide or a nonpolynucleotide sequence) is subjected to a

source of exonuclease activity. Serviceable sources of exonuclease activity may be an enzyme with 3' exonuclease activity, an enzyme with 5' exonuclease activity, an enzyme with both 3' exonuclease activity and 5' exonuclease activity, and any combination thereof. An exonuclease can be used to liberate nucleotides from one or both ends of a
5 linear double stranded polynucleotide, and from one to all ends of a branched polynucleotide having more than two ends. The mechanism of action of this liberation is believed to be comprised of an enzymatically-catalyzed hydrolysis of terminal nucleotides, and can be allowed to proceed in a time-dependent fashion, allowing experimental control of the progression of the enzymatic process.

10

By contrast, a non-enzymatic step may be used to shuffle, assemble, reassemble, recombine, and/or concatenate polynucleotide building blocks that is comprised of subjecting a working sample to denaturing (or "melting") conditions (for example, by changing temperature, pH, and /or salinity conditions) so as to melt a working set of
15 double stranded polynucleotides into single polynucleotide strands. For shuffling, it is desirable that the single polynucleotide strands participate to some extent in annealment with different hybridization partners (i.e. and not merely revert to exclusive reannealment between what were former partners before the denaturation step). The presence of the former hybridization partners in the reaction vessel, however, does not preclude, and may
20 sometimes even favor, reannealment of a single stranded polynucleotide with its former partner, to recreate an original double stranded polynucleotide.

In contrast to this non-enzymatic shuffling step comprised of subjecting double stranded polynucleotide building blocks to denaturation, followed by annealment, the
25 instant invention further provides an exonuclease-based approach requiring no denaturation – rather, the avoidance of denaturing conditions and the maintenance of double stranded polynucleotide substrates in annealed (i.e. non-denatured) state are necessary conditions for the action of exonucleases (e.g., exonuclease III and red alpha gene product). Additionally in contrast, the generation of single stranded polynucleotide

sequences capable of hybridizing to other single stranded polynucleotide sequences is the result of covalent cleavage – and hence sequence destruction - in one of the hybridization partners. For example, an exonuclease III enzyme may be used to enzymatically liberate 3' terminal nucleotides in one hybridization strand (to achieve covalent hydrolysis in that polynucleotide strand); and this favors hybridization of the remaining single strand to a new partner (since its former partner was subjected to covalent cleavage).

By way of further illustration, a specific exonuclease, namely exonuclease III is provided herein as an example of a 3' exonuclease; however, other exonucleases may also be used, including enzymes with 5' exonuclease activity and enzymes with 3' exonuclease activity, and including enzymes not yet discovered and enzymes not yet developed. It is particularly appreciated that enzymes can be discovered, optimized (e.g. engineered by directed evolution), or both discovered and optimized specifically for the instantly disclosed approach that have more optimal rates &/or more highly specific activities &/or greater lack of unwanted activities. In fact it is expected that the instant invention may encourage the discovery &/or development of such designer enzymes. In sum, this invention may be practiced with a variety of currently available exonuclease enzymes, as well enzymes not yet discovered and enzymes not yet developed.

The exonuclease action of exonuclease III requires a working double stranded polynucleotide end that is either blunt or has a 5' overhang, and the exonuclease action is comprised of enzymatically liberating 3' terminal nucleotides, leaving a single stranded 5' end that becomes longer and longer as the exonuclease action proceeds (see Figure 1). Any 5' overhangs produced by this approach may be used to hybridize to another single stranded polynucleotide sequence (which may also be a single stranded polynucleotide or a terminal overhang of a partially double stranded polynucleotide) that shares enough homology to allow hybridization. The ability of these exonuclease III-generated single stranded sequences (e.g. in 5' overhangs) to hybridize to other single stranded sequences

allows two or more polynucleotides to be shuffled, assembled, reassembled, &/or concatenated.

Furthermore, it is appreciated that one can protect the end of a double stranded polynucleotide or render it susceptible to a desired enzymatic action of a serviceable exonuclease as necessary. For example, a double stranded polynucleotide end having a 3' overhang is not susceptible to the exonuclease action of exonuclease III. However, it may be rendered susceptible to the exonuclease action of exonuclease III by a variety of means; for example, it may be blunted by treatment with a polymerase, cleaved to provide a blunt end or a 5' overhang, joined (ligated or hybridized) to another double stranded polynucleotide to provide a blunt end or a 5' overhang, hybridized to a single stranded polynucleotide to provide a blunt end or a 5' overhang, or modified by any of a variety of means).

According to one aspect, an exonuclease may be allowed to act on one or on both ends of a linear double stranded polynucleotide and proceed to completion, to near completion, or to partial completion. When the exonuclease action is allowed to go to completion, the result will be that the length of each 5' overhang will be extend far towards the middle region of the polynucleotide in the direction of what might be considered a "rendezvous point" (which may be somewhere near the polynucleotide midpoint). Ultimately, this results in the production of single stranded polynucleotides (that can become dissociated) that are each about half the length of the original double stranded polynucleotide (see Figure 1). Alternatively, an exonuclease-mediated reaction can be terminated before proceeding to completion.

Thus this exonuclease-mediated approach is serviceable for shuffling, assembling &/or reassembling, recombining, and concatenating polynucleotide building blocks, which polynucleotide building blocks can be up to ten bases long or tens of bases long or

hundreds of bases long or thousands of bases long or tens of thousands of bases long or hundreds of thousands of bases long or millions of bases long or even longer.

This exonuclease-mediated approach is based on the action of double stranded
5 DNA specific exodeoxyribonuclease activity of *E. coli* exonuclease III. Substrates for exonuclease III may be generated by subjecting a double stranded polynucleotide to fragmentation. Fragmentation may be achieved by mechanical means (e.g., shearing, sonication, etc.), by enzymatic means (e.g. using restriction enzymes), and by any combination thereof. Fragments of a larger polynucleotide may also be generated by
10 polymerase-mediated synthesis.

Exonuclease III is a 28K monomeric enzyme, product of the *xthA* gene of *E. coli* with four known activities: exodeoxyribonuclease (alternatively referred to as exonuclease herein), RNaseH, DNA-3'-phosphatase, and AP endonuclease. The
15 exodeoxyribonuclease activity is specific for double stranded DNA. The mechanism of action is thought to involve enzymatic hydrolysis of DNA from a 3' end progressively towards a 5' direction, with formation of nucleoside 5'-phosphates and a residual single strand. The enzyme does not display efficient hydrolysis of single stranded DNA, single-stranded RNA, or double-stranded RNA; however it degrades RNA in an DNA-RNA
20 hybrid releasing nucleoside 5'-phosphates. The enzyme also releases inorganic phosphate specifically from 3'phosphomonoester groups on DNA, but not from RNA or short oligonucleotides. Removal of these groups converts the terminus into a primer for DNA polymerase action.

25 Additional examples of enzymes with exonuclease activity include red-alpha and venom phosphodiesterases. Red alpha (*redα*) gene product (also referred to as lambda exonuclease) is of bacteriophage λ origin. The *redα* gene is transcribed from the leftward promoter and its product is involved (24 kD) in recombination. Red alpha gene product acts processively from 5'-phosphorylated termini to liberate mononucleotides from

duplex DNA (Takahashi & Kobayashi, 1990). Venom phosphodiesterases (Laskowski, 1980) is capable of rapidly opening supercoiled DNA.

5 2.11.2.3. NON-STOCHASTIC LIGATION REASSEMBLY

In one aspect, the present invention provides a non-stochastic method termed synthetic ligation reassembly (SLR), that is somewhat related to stochastic shuffling, save that the nucleic acid building blocks are not shuffled or concatenated or chimerized
10 randomly, but rather are assembled non-stochastically.

A particularly glaring difference is that the instant SLR method does not depend on the presence of a high level of homology between polynucleotides to be shuffled. In contrast, prior methods, particularly prior stochastic shuffling methods require that
15 presence of a high level of homology, particularly at coupling sites, between polynucleotides to be shuffled. Accordingly these prior methods favor the regeneration of the original progenitor molecules, and are suboptimal for generating large numbers of novel progeny chimeras, particularly full-length progenies. The instant invention, on the other hand, can be used to non-stochastically generate libraries (or sets) of progeny
20 molecules comprised of over 10^{100} different chimeras. Conceivably, SLR can even be used to generate libraries comprised of over 10^{1000} different progeny chimeras with (no upper limit in sight).

Thus, in one aspect, the present invention provides a method, which method is
25 non-stochastic, of producing a set of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design, which method is comprised of the steps of generating by design a plurality of specific nucleic acid building blocks having serviceable mutually compatible ligatable ends, and assembling these nucleic acid building blocks, such that a designed overall assembly order is achieved.

The mutually compatible ligatable ends of the nucleic acid building blocks to be assembled are considered to be "serviceable" for this type of ordered assembly if they enable the building blocks to be coupled in predetermined orders. Thus, in one aspect, 5 the overall assembly order in which the nucleic acid building blocks can be coupled is specified by the design of the ligatable ends and, if more than one assembly step is to be used, then the overall assembly order in which the nucleic acid building blocks can be coupled is also specified by the sequential order of the assembly step(s). Figure 4, Panel C illustrates an exemplary assembly process comprised of 2 sequential steps to achieve a 10 designed (non-stochastic) overall assembly order for five nucleic acid building blocks. In a preferred embodiment of this invention, the annealed building pieces are treated with an enzyme, such as a ligase (e.g. T4 DNA ligase), achieve covalent bonding of the building pieces.

15 In a preferred embodiment, the design of nucleic acid building blocks is obtained upon analysis of the sequences of a set of progenitor nucleic acid templates that serve as a basis for producing a progeny set of finalized chimeric nucleic acid molecules. These progenitor nucleic acid templates thus serve as a source of sequence information that aids in the design of the nucleic acid building blocks that are to be mutagenized, i.e. 20 chimerized or shuffled.

In one exemplification, this invention provides for the chimerization of a family of related genes and their encoded family of related products. In a particular exemplification, the encoded products are enzymes. As a representative list of families 25 of enzymes which may be mutagenized in accordance with the aspects of the present invention, there may be mentioned, the following enzymes and their functions:

- 1 **Lipase/Esterase**
- a. Enantioselective hydrolysis of esters (lipids)/ thioesters
 - 1) Resolution of racemic mixtures
 - 2) Synthesis of optically active acids or alcohols from *meso*-diesters
 - 5 b. Selective syntheses
 - 1) Regiospecific hydrolysis of carbohydrate esters
 - 2) Selective hydrolysis of cyclic secondary alcohols
 - c. Synthesis of optically active esters, lactones, acids, alcohols
 - 1) Transesterification of activated/nonactivated esters
 - 10 2) Interesterification
 - 3) Optically active lactones from hydroxyesters
 - 4) Regio- and enantioselective ring opening of anhydrides
 - d. Detergents
 - e. Fat/Oil conversion
 - 15 f. Cheese ripening
- 2 **Protease**
- a. Ester/amide synthesis
 - b. Peptide synthesis
 - 20 c. Resolution of racemic mixtures of amino acid esters
 - d. Synthesis of non-natural amino acids
 - e. Detergents/protein hydrolysis
- 3 **Glycosidase/Glycosyl transferase**
- 25 a. Sugar/polymer synthesis
 - b. Cleavage of glycosidic linkages to form mono, di-and oligosaccharides
 - c. Synthesis of complex oligosaccharides
 - d. Glycoside synthesis using UDP-galactosyl transferase
 - e. Transglycosylation of disaccharides, glycosyl fluorides, aryl galactosides

- f. Glycosyl transfer in oligosaccharide synthesis
 - g. Diastereoselective cleavage of β -glucosylsulfoxides
 - h. Asymmetric glycosylations
 - i. Food processing
 - 5 j. Paper processing
- 4 **Phosphatase/Kinase**
- a. Synthesis/hydrolysis of phosphate esters
 - 1) Regio-, enantioselective phosphorylation
 - 10 2) Introduction of phosphate esters
 - 3) Synthesize phospholipid precursors
 - 4) Controlled polynucleotide synthesis
 - b. Activate biological molecule
 - c. Selective phosphate bond formation without protecting groups
- 15
- 5 **Mono/Dioxygenase**
- a. Direct oxyfunctionalization of unactivated organic substrates
 - b. Hydroxylation of alkane, aromatics, steroids
 - c. Epoxidation of alkenes
 - 20 d. Enantioselective sulphoxidation
 - e. Regio- and stereoselective Bayer-Villiger oxidations
- 6 **Haloperoxidase**
- a. Oxidative addition of halide ion to nucleophilic sites
 - 25 b. Addition of hypohalous acids to olefinic bonds
 - c. Ring cleavage of cyclopropanes
 - d. Activated aromatic substrates converted to *ortho* and *para* derivatives
 - e. 1,3 diketones converted to 2-halo-derivatives

- f. Heteroatom oxidation of sulfur and nitrogen containing substrates
 - g. Oxidation of enol acetates, alkynes and activated aromatic rings
- 7 **Lignin peroxidase/Diarylpropane peroxidase**
- 5 a. Oxidative cleavage of C-C bonds
 - b. Oxidation of benzylic alcohols to aldehydes
 - c. Hydroxylation of benzylic carbons
 - d. Phenol dimerization
 - e. Hydroxylation of double bonds to form diols
 - 10 f. Cleavage of lignin aldehydes
- 8 **Epoxide hydrolase**
- a. Synthesis of enantiomerically pure bioactive compounds
 - b. Regio- and enantioselective hydrolysis of epoxide
 - 15 c. Aromatic and olefinic epoxidation by monooxygenases to form epoxides
 - d. Resolution of racemic epoxides
 - e. Hydrolysis of steroid epoxides
- 9 **Nitrile hydratase/nitrilase**
- 20 a. Hydrolysis of aliphatic nitriles to carboxamides
 - b. Hydrolysis of aromatic, heterocyclic, unsaturated aliphatic nitriles to
 corresponding acids
 - c. Hydrolysis of acrylonitrile
 - d. Production of aromatic and carboxamides, carboxylic acids (nicotinamide,
25 picolinamide, isonicotinamide)
 - e. Regioselective hydrolysis of acrylic dinitrile
 - f. α -amino acids from α -hydroxynitriles

10 **Transaminase**

- a. Transfer of amino groups into oxo-acids

5 11 **Amidase/Acylase**

- a. Hydrolysis of amides, amidines, and other C-N bonds
b. Non-natural amino acid resolution and synthesis

These exemplifications, while illustrating certain specific aspects of the invention,
10 do not portray the limitations or circumscribe the scope of the disclosed invention.

Thus according to one aspect of this invention, the sequences of a plurality of progenitor nucleic acid templates are aligned in order to select one or more demarcation points, which demarcation points can be located at an area of homology, and are
15 comprised of one or more nucleotides, and which demarcation points are shared by at least two of the progenitor templates. The demarcation points can be used to delineate the boundaries of nucleic acid building blocks to be generated. Thus, the demarcation points identified and selected in the progenitor molecules serve as potential chimerization points in the assembly of the progeny molecules.

20

Preferably a serviceable demarcation point is an area of homology (comprised of at least one homologous nucleotide base) shared by at least two progenitor templates. More preferably a serviceable demarcation point is an area of homology that is shared by at least half of the progenitor templates. More preferably still a serviceable demarcation
25 point is an area of homology that is shared by at least two thirds of the progenitor templates. Even more preferably a serviceable demarcation points is an area of homology that is shared by at least three fourths of the progenitor templates. Even more preferably still a serviceable demarcation points is an area of homology that is shared by

at almost all of the progenitor templates. Even more preferably still a serviceable demarcation point is an area of homology that is shared by all of the progenitor templates.

5 The process of designing nucleic acid building blocks and of designing the mutually compatible ligatable ends of the nucleic acid building blocks to be assembled is illustrated in Figures 6 and 7. As shown, the alignment of a set of progenitor templates reveals several naturally occurring demarcation points, and the identification of demarcation points shared by these templates helps to non-stochastically determine the building blocks to be generated and used for the generation of the progeny chimeric
10 molecules.

In a preferred embodiment, this invention provides that the ligation reassembly process is performed exhaustively in order to generate an exhaustive library. In other words, all possible ordered combinations of the nucleic acid building blocks are
15 represented in the set of finalized chimeric nucleic acid molecules. At the same time, in a particularly preferred embodiment, the assembly order (i.e. the order of assembly of each building block in the 5' to 3' sequence of each finalized chimeric nucleic acid) in each combination is by design (or non-stochastic). Because of the non-stochastic nature of this invention, the possibility of unwanted side products is greatly reduced.

20

In another preferred embodiment, this invention provides that, the ligation reassembly process is performed systematically, for example in order to generate a systematically compartmentalized library, with compartments that can be screened systematically, e.g. one by one. In other words this invention provides that, through the
25 selective and judicious use of specific nucleic acid building blocks, coupled with the selective and judicious use of sequentially stepped assembly reactions, an experimental design can be achieved where specific sets of progeny products are made in each of several reaction vessels. This allows a systematic examination and screening procedure

to be performed. Thus, it allows a potentially very large number of progeny molecules to be examined systematically in smaller groups.

Because of its ability to perform chimerizations in a manner that is highly flexible
5 yet exhaustive and systematic as well, particularly when there is a low level of homology
among the progenitor molecules, the instant invention provides for the generation of a
library (or set) comprised of a large number of progeny molecules. Because of the non-
stochastic nature of the instant ligation reassembly invention, the progeny molecules
generated preferably comprise a library of finalized chimeric nucleic acid molecules
10 having an overall assembly order that is chosen by design. In a particularly preferred
embodiment of this invention, such a generated library is comprised of preferably greater
than 10^3 different progeny molecular species, more preferably greater than 10^5 different
progeny molecular species, more preferably still greater than 10^{10} different progeny
molecular species, more preferably still greater than 10^{15} different progeny molecular
15 species, more preferably still greater than 10^{20} different progeny molecular species, more
preferably still greater than 10^{30} different progeny molecular species, more preferably
still greater than 10^{40} different progeny molecular species, more preferably still greater
than 10^{50} different progeny molecular species, more preferably still greater than 10^{60}
different progeny molecular species, more preferably still greater than 10^{70} different
20 progeny molecular species, more preferably still greater than 10^{80} different progeny
molecular species, more preferably still greater than 10^{100} different progeny molecular
species, more preferably still greater than 10^{110} different progeny molecular species, more
preferably still greater than 10^{120} different progeny molecular species, more preferably
still greater than 10^{130} different progeny molecular species, more preferably still greater
25 than 10^{140} different progeny molecular species, more preferably still greater than 10^{150}
different progeny molecular species, more preferably still greater than 10^{175} different
progeny molecular species, more preferably still greater than 10^{200} different progeny
molecular species, more preferably still greater than 10^{300} different progeny molecular
species, more preferably still greater than 10^{400} different progeny molecular species, more

preferably still greater than 10^{500} different progeny molecular species, and even more preferably still greater than 10^{1000} different progeny molecular species.

In one aspect, a set of finalized chimeric nucleic acid molecules, produced as
5 described is comprised of a polynucleotide encoding a polypeptide. According to one preferred embodiment, this polynucleotide is a gene, which may be a man-made gene. According to another preferred embodiment, this polynucleotide is a gene pathway, which may be a man-made gene pathway. This invention provides that one or more man-made genes generated by this invention may be incorporated into a man-made gene
10 pathway, such as pathway operable in a eukaryotic organism (including a plant).

It is appreciated that the power of this invention is exceptional, as there is much freedom of choice and control regarding the selection of demarcation points, the size and number of the nucleic acid building blocks, and the size and design of the couplings. It is
15 appreciated, furthermore, that the requirement for intermolecular homology is highly relaxed for the operability of this invention. In fact, demarcation points can even be chosen in areas of little or no intermolecular homology. For example, because of codon wobble, i.e. the degeneracy of codons, nucleotide substitutions can be introduced into nucleic acid building blocks without altering the amino acid originally encoded in the
20 corresponding progenitor template. Alternatively, a codon can be altered such that the coding for an originally amino acid is altered. This invention provides that such substitutions can be introduced into the nucleic acid building block in order to increase the incidence of intermolecularly homologous demarcation points and thus to allow an increased number of couplings to be achieved among the building blocks, which in turn
25 allows a greater number of progeny chimeric molecules to be generated.

In another exemplification, the synthetic nature of the step in which the building blocks are generated allows the design and introduction of nucleotides (e.g. one or more nucleotides, which may be, for example, codons or introns or regulatory sequences) that

can later be optionally removed in an in vitro process (e.g. by mutagenesis) or in an in vivo process (e.g. by utilizing the gene splicing ability of a host organism). It is appreciated that in many instances the introduction of these nucleotides may also be desirable for many other reasons in addition to the potential benefit of creating a serviceable demarcation point.

Thus, according to another embodiment, this invention provides that a nucleic acid building block can be used to introduce an intron. Thus, this invention provides that functional introns may be introduced into a man-made gene of this invention. This invention also provides that functional introns may be introduced into a man-made gene pathway of this invention. Accordingly, this invention provides for the generation of a chimeric polynucleotide that is a man-made gene containing one (or more) artificially introduced intron(s).

Accordingly, this invention also provides for the generation of a chimeric polynucleotide that is a man-made gene pathway containing one (or more) artificially introduced intron(s). Preferably, the artificially introduced intron(s) are functional in one or more host cells for gene splicing much in the way that naturally-occurring introns serve functionally in gene splicing. This invention provides a process of producing man-made intron-containing polynucleotides to be introduced into host organisms for recombination and/or splicing.

The ability to achieve chimerizations, using couplings as described herein, in areas of little or no homology among the progenitor molecules, is particularly useful, and in fact critical, for the assembly of novel gene pathways. This invention thus provides for the generation of novel man-made gene pathways using synthetic ligation reassembly. In a particular aspect, this is achieved by the introduction of regulatory sequences, such as promoters, that are operable in an intended host, to confer operability to a novel gene pathway when it is introduced into the intended host. In a particular exemplification, this

invention provides for the generation of novel man-made gene pathways that is operable in a plurality of intended hosts (e.g. in a microbial organism as well as in a plant cell). This can be achieved, for example, by the introduction of a plurality of regulatory sequences, comprised of a regulatory sequence that is operable in a first intended host
5 and a regulatory sequence that is operable in a second intended host. A similar process can be performed to achieve operability of a gene pathway in a third intended host species, etc. The number of intended host species can be each integer from 1 to 10 or alternatively over 10. Alternatively, for example, operability of a gene pathway in a plurality of intended hosts can be achieved by the introduction of a regulatory sequence
10 having intrinsic operability in a plurality of intended hosts.

Thus, according to a particular embodiment, this invention provides that a nucleic acid building block can be used to introduce a regulatory sequence, particularly a regulatory sequence for gene expression. Preferred regulatory sequences include, but are
15 not limited to, those that are man-made, and those found in archeal, bacterial, eukaryotic (including mitochondrial), viral, and prionic or prion-like organisms. Preferred regulatory sequences include but are not limited to, promoters, operators, and activator binding sites. Thus, this invention provides that functional regulatory sequences may be introduced into a man-made gene of this invention. This invention also provides that
20 functional regulatory sequences may be introduced into a man-made gene pathway of this invention.

Accordingly, this invention provides for the generation of a chimeric polynucleotide that is a man-made gene containing one (or more) artificially introduced
25 regulatory sequence(s). Accordingly, this invention also provides for the generation of a chimeric polynucleotide that is a man-made gene pathway containing one (or more) artificially introduced regulatory sequence(s). Preferably, an artificially introduced regulatory sequence(s) is operatively linked to one or more genes in the man-made polynucleotide, and are functional in one or more host cells.

Preferred bacterial promoters that are serviceable for this invention include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Serviceable eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Particular plant regulatory sequences include promoters active in directing transcription in plants, either constitutively or stage and/or tissue specific, depending on the use of the plant or parts thereof. These promoters include, but are not limited to promoters showing constitutive expression, such as the 35S promoter of Cauliflower Mosaic Virus (CaMV) (Guilley et al., 1982), those for leaf-specific expression, such as the promoter of the ribulose biphosphate carboxylase small subunit gene (Coruzzi et al., 1984), those for root-specific expression, such as the promoter from the glutamin synthase gene (Tingey et al., 1987), those for seed-specific expression, such as the cruciferin A promoter from *Brassica napus* (Ryan et al., 1989), those for tuber-specific expression, such as the class-I patatin promoter from potato (Rocha-Sasa et al., 1989; Wenzler et al., 1989) or those for fruit-specific expression, such as the polygalacturonase (PG) promoter from tomato (Bird et al., 1988).

Other regulatory sequences that are preferred for this invention include terminator sequences and polyadenylation signals and any such sequence functioning as such in plants, the choice of which is within the level of the skilled artisan. An example of such sequences is the 3' flanking region of the nopaline synthase (nos) gene of *Agrobacterium tumefaciens* (Bevan, 1984). The regulatory sequences may also include enhancer sequences, such as found in the 35S promoter of CaMV, and mRNA stabilizing sequences such as the leader sequence of Alfalfa Mosaic Virus (AIMV) RNA4 (Brederode et al., 1980) or any other sequences functioning in a like manner.

A man-made genes produced using this invention can also serve as a substrate for recombination with another nucleic acid. Likewise, a man-made gene pathway produced using this invention can also serve as a substrate for recombination with another nucleic

acid. In a preferred instance, the recombination is facilitated by, or occurs at, areas of homology between the man-made intron-containing gene and a nucleic acid which serves as a recombination partner. In a particularly preferred instance, the recombination partner may also be a nucleic acid generated by this invention, including a man-made gene or a man-made gene pathway. Recombination may be facilitated by or may occur at areas of homology that exist at the one (or more) artificially introduced intron(s) in the man-made gene.

The synthetic ligation reassembly method of this invention utilizes a plurality of nucleic acid building blocks, each of which preferably has two ligatable ends. The two ligatable ends on each nucleic acid building block may be two blunt ends (i.e. each having an overhang of zero nucleotides), or preferably one blunt end and one overhang, or more preferably still two overhangs.

A serviceable overhang for this purpose may be a 3' overhang or a 5' overhang. Thus, a nucleic acid building block may have a 3' overhang or alternatively a 5' overhang or alternatively two 3' overhangs or alternatively two 5' overhangs. The overall order in which the nucleic acid building blocks are assembled to form a finalized chimeric nucleic acid molecule is determined by purposeful experimental design and is not random.

According to one preferred embodiment, a nucleic acid building block is generated by chemical synthesis of two single-stranded nucleic acids (also referred to as single-stranded oligos) and contacting them so as to allow them to anneal to form a double-stranded nucleic acid building block.

A double-stranded nucleic acid building block can be of variable size. The sizes of these building blocks can be small or large depending on the choice of the experimenter. Preferred sizes for building block range from 1 base pair (not including any overhangs) to 100,000 base pairs (not including any overhangs). Other preferred size

ranges are also provided, which have lower limits of from 1 bp to 10,000 bp (including every integer value in between), and upper limits of from 2 bp to 100, 000 bp (including every integer value in between).

5 It is appreciated that current methods of polymerase-based amplification can be used to generate double-stranded nucleic acids of up to thousands of base pairs, if not tens of thousands of base pairs, in length with high fidelity. Chemical synthesis (e.g. phosphoramidite-based) can be used to generate nucleic acids of up to hundreds of nucleotides in length with high fidelity; however, these can be assembled, e.g. using
10 overhangs or sticky ends, to form double-stranded nucleic acids of up to thousands of base pairs, if not tens of thousands of base pairs, in length if so desired.

 A combination of methods (e.g. phosphoramidite-based chemical synthesis and PCR) can also be used according to this invention. Thus, nucleic acid building block
15 made by different methods can also be used in combination to generate a progeny molecule of this invention.

 The use of chemical synthesis to generate nucleic acid building blocks is particularly preferred in this invention & is advantageous for other reasons as well,
20 including procedural safety and ease. No cloning or harvesting or actual handling of any biological samples is required. The design of the nucleic acid building blocks can be accomplished on paper. Accordingly, this invention teaches an advance in procedural safety in recombinant technologies.

25 Nonetheless, according to one preferred embodiment, a double-stranded nucleic acid building block according to this invention may also be generated by polymerase-based amplification of a polynucleotide template. In a non-limiting exemplification, as illustrated in Figure 2, a first polymerase-based amplification reaction using a first set of primers, F₂ and R₁, is used to generate a blunt-ended product (labeled Reaction 1, Product

1), which is essentially identical to Product A. A second polymerase-based amplification reaction using a second set of primers, F_1 and R_2 , is used to generate a blunt-ended product (labeled Reaction 2, Product 2), which is essentially identical to Product B. These two products are mixed and allowed to melt and anneal, generating potentially
5 useful double-stranded nucleic acid building blocks with two overhangs. In the example of Fig. 2, the product with the 3' overhangs (Product C) is selected by nuclease-based degradation of the other 3 products using a 3' acting exonuclease, such as exonuclease III. It is appreciated that a 5' acting exonuclease (e.g. red alpha) may be also be used, for example to select Product D instead. It is also appreciated that other selection means can
10 also be used, including hybridization-based means, and that these means can incorporate a further means, such as a magnetic bead-based means, to facilitate separation of the desired product.

Many other methods exist by which a double-stranded nucleic acid building block
15 can be generated that is serviceable for this invention; and these are known in the art and can be readily performed by the skilled artisan.

According to particularly preferred embodiment, a double-stranded nucleic acid building block that is serviceable for this invention is generated by first generating two
20 single stranded nucleic acids and allowing them to anneal to form a double-stranded nucleic acid building block. The two strands of a double-stranded nucleic acid building block may be complementary at every nucleotide apart from any that form an overhang; thus containing no mismatches, apart from any overhang(s). According to another embodiment, the two strands of a double-stranded nucleic acid building block are
25 complementary at fewer than every nucleotide apart from any that form an overhang. Thus, according to this embodiment, a double-stranded nucleic acid building block can be used to introduce codon degeneracy. Preferably the codon degeneracy is introduced using the site-saturation mutagenesis described herein, using one or more N,N,G/T cassettes or alternatively using one or more N,N,N cassettes.

Contained within an exemplary experimental design for achieving an ordered assembly according to this invention are:

- 5 1) The design of specific nucleic acid building blocks.
 2) The design of specific ligatable ends on each nucleic acid building block.
 3) The design of a particular order of assembly of the nucleic acid building blocks.

10 An overhang may be a 3' overhang or a 5' overhang. An overhang may also have a terminal phosphate group or alternatively may be devoid of a terminal phosphate group (having, e.g., a hydroxyl group instead). An overhang may be comprised of any number of nucleotides. Preferably an overhang is comprised of 0 nucleotides (as in a blunt end) to 10,000 nucleotides. Thus, a wide range of overhang sizes may be serviceable.

15 Accordingly, the lower limit may be each integer from 1-200 and the upper limit may be each integer from 2-10,000. According to a particular exemplification, an overhang may consist of anywhere from 1 nucleotide to 200 nucleotides (including every integer value in between).

20 The final chimeric nucleic acid molecule may be generated by sequentially assembling 2 or more building blocks at a time until all the designated building blocks have been assembled. A working sample may optionally be subjected to a process for size selection or purification or other selection or enrichment process between the performance of two assembly steps. Alternatively, the final chimeric nucleic acid
25 molecule may be generated by assembling all the designated building blocks at once in one step.

Utility

The *in vivo* recombination method of this invention can be performed blindly on a pool of unknown hybrids or alleles of a specific polynucleotide or sequence. However, it is not necessary to know the actual DNA or RNA sequence of the specific
5 polynucleotide.

The approach of using recombination within a mixed population of genes can be useful for the generation of any useful proteins, for example, interleukin I, antibodies, tPA and growth hormone. This approach may be used to generate proteins having altered
10 specificity or activity. The approach may also be useful for the generation of hybrid nucleic acid sequences, for example, promoter regions, introns, exons, enhancer sequences, 3' untranslated regions or 5' untranslated regions of genes. Thus this approach may be used to generate genes having increased rates of expression. This approach may also be useful in the study of repetitive DNA sequences. Finally, this
15 approach may be useful to mutate ribozymes or aptamers.

Scaffold-like regions separating regions of diversity in proteins may be particularly suitable for the methods of this invention. The conserved scaffold determines the overall folding by self-association, while displaying relatively unrestricted
20 loops that mediate the specific binding. Examples of such scaffolds are the immunoglobulin beta barrel, and the four-helix bundle. The methods of this invention can be used to create scaffold-like proteins with various combinations of mutated sequences for binding.

25 The equivalents of some standard genetic matings may also be performed by the methods of this invention. For example, a "molecular" backcross can be performed by repeated mixing of the hybrid's nucleic acid with the wild-type nucleic acid while selecting for the mutations of interest. As in traditional breeding, this approach can be used to combine phenotypes from different sources into a background of choice. It is

useful, for example, for the removal of neutral mutations that affect unselected characteristics (i.e. immunogenicity). Thus it can be useful to determine which mutations in a protein are involved in the enhanced biological activity and which are not.

5

2.11.2.4. END-SELECTION

This invention provides a method for selecting a subset of polynucleotides from a starting set of polynucleotides, which method is based on the ability to discriminate one or more selectable features (or selection markers) present anywhere in a working polynucleotide, so as to allow one to perform selection for (positive selection) &/or against (negative selection) each selectable polynucleotide. In a preferred aspect, a method is provided termed end-selection, which method is based on the use of a selection marker located in part or entirely in a terminal region of a selectable polynucleotide, and such a selection marker may be termed an "end-selection marker".

End-selection may be based on detection of naturally occurring sequences or on detection of sequences introduced experimentally (including by any mutagenesis procedure mentioned herein and not mentioned herein) or on both, even within the same polynucleotide. An end-selection marker can be a structural selection marker or a functional selection marker or both a structural and a functional selection marker. An end-selection marker may be comprised of a polynucleotide sequence or of a polypeptide sequence or of any chemical structure or of any biological or biochemical tag, including markers that can be selected using methods based on the detection of radioactivity, of enzymatic activity, of fluorescence, of any optical feature, of a magnetic property (e.g. using magnetic beads), of immunoreactivity, and of hybridization.

End-selection may be applied in combination with any method serviceable for performing mutagenesis. Such mutagenesis methods include, but are not limited to,

methods described herein (supra and infra). Such methods include, by way of non-limiting exemplification, any method that may be referred herein or by others in the art by any of the following terms: "saturation mutagenesis", "shuffling", "recombination", "re-assembly", "error-prone PCR", "assembly PCR", "sexual PCR", "crossover PCR",
5 "oligonucleotide primer-directed mutagenesis", "recursive (&/or exponential) ensemble mutagenesis (see **Arkin** and **Youvan**, 1992)", "cassette mutagenesis", "in vivo mutagenesis", and "in vitro mutagenesis". Moreover, end-selection may be performed on molecules produced by any mutagenesis &/or amplification method (see, e.g., **Arnold**, 1993; **Caldwell** and **Joyce**, 1992; **Stemmer**, 1994; following which method it is desirable
10 to select for (including to screen for the presence of) desirable progeny molecules.

In addition, end-selection may be applied to a polynucleotide apart from any mutagenesis method. In a preferred embodiment, end-selection, as provided herein, can be used in order to facilitate a cloning step, such as a step of ligation to another
15 polynucleotide (including ligation to a vector). This invention thus provides for end-selection as a serviceable means to facilitate library construction, selection &/or enrichment for desirable polynucleotides, and cloning in general.

In a particularly preferred embodiment, end-selection can be based on (positive)
20 selection for a polynucleotide; alternatively end-selection can be based on (negative) selection against a polynucleotide; and alternatively still, end-selection can be based on both (positive) selection for, and on (negative) selection against, a polynucleotide. End-selection, along with other methods of selection &/or screening, can be performed in an iterative fashion, with any combination of like or unlike selection &/or screening
25 methods and serviceable mutagenesis methods, all of which can be performed in an iterative fashion and in any order, combination, and permutation.

It is also appreciated that, according to one embodiment of this invention, end-selection may also be used to select a polynucleotide is at least in part: circular (e.g. a

plasmid or any other circular vector or any other polynucleotide that is partly circular), &/or branched, &/or modified or substituted with any chemical group or moiety. In accord with this embodiment, a polynucleotide may be a circular molecule comprised of an intermediate or central region, which region is flanked on a 5' side by a 5' flanking
5 region (which, for the purpose of end-selection, serves in like manner to a 5' terminal region of a non-circular polynucleotide) and on a 3' side by a 3' terminal region (which, for the purpose of end-selection, serves in like manner to a 3' terminal region of a non-circular polynucleotide). As used in this non-limiting exemplification, there may be sequence overlap between any two regions or even among all three regions.

10

In one non-limiting aspect of this invention, end-selection of a linear polynucleotide is performed using a general approach based on the presence of at least one end-selection marker located at or near a polynucleotide end or terminus (that can be either a 5' end or a 3' end). In one particular non-limiting exemplification, end-selection
15 is based on selection for a specific sequence at or near a terminus such as, but not limited to, a sequence recognized by an enzyme that recognizes a polynucleotide sequence. An enzyme that recognizes and catalyzes a chemical modification of a polynucleotide is referred to herein as a polynucleotide-acting enzyme. In a preferred embodiment, serviceable polynucleotide-acting enzymes are exemplified non-exclusively by enzymes
20 with polynucleotide-cleaving activity, enzymes with polynucleotide-methylating activity, enzymes with polynucleotide-ligating activity, and enzymes with a plurality of distinguishable enzymatic activities (including non-exclusively, e.g., both polynucleotide-cleaving activity and polynucleotide-ligating activity).

25 Relevant polynucleotide-acting enzymes thus also include any commercially available or non-commercially available polynucleotide endonucleases and their companion methylases including those catalogued at the website <http://www.neb.com/rebase>, and those mentioned in the following cited reference (Roberts and Macelis, 1996). Preferred polynucleotide endonucleases include – but are

not limited to – type II restriction enzymes (including type IIS), and include enzymes that cleave both strands of a double stranded polynucleotide (e.g. *Not* I, which cleaves both strands at 5'...GC/GGCCGC...3') and enzymes that cleave only one strand of a double stranded polynucleotide, i.e. enzymes that have polynucleotide-nicking activity, (e.g. N. *Bst* I, which cleaves only one strand at 5'...GAGTCNNNN/N...3'). Relevant polynucleotide-acting enzymes also include type III restriction enzymes.

It is appreciated that relevant polynucleotide-acting enzymes also include any enzymes that may be developed in the future, though currently unavailable, that are serviceable for generating a ligation compatible end, preferably a sticky end, in a polynucleotide.

In one preferred exemplification, a serviceable selection marker is a restriction site in a polynucleotide that allows a corresponding type II (or type IIS) restriction enzyme to cleave an end of the polynucleotide so as to provide a ligatable end (including a blunt end or alternatively a sticky end with at least a one base overhang) that is serviceable for a desirable ligation reaction without cleaving the polynucleotide internally in a manner that destroys a desired internal sequence in the polynucleotide. Thus it is provided that, among relevant restriction sites, those sites that do not occur internally (i.e. that do not occur apart from the termini) in a specific working polynucleotide are preferred when the use of a corresponding restriction enzyme(s) is not intended to cut the working polynucleotide internally. This allows one to perform restriction digestion reactions to completion or to near completion without incurring unwanted internal cleavage in a working polynucleotide.

According to a preferred aspect, it is thus preferable to use restriction sites that are not contained, or alternatively that are not expected to be contained, or alternatively that unlikely to be contained (e.g. when sequence information regarding a working polynucleotide is incomplete) internally in a polynucleotide to be subjected to end-

selection. In accordance with this aspect, it is appreciated that restriction sites that occur relatively infrequently are usually preferred over those that occur more frequently. On the other hand it is also appreciated that there are occasions where internal cleavage of a polypeptide is desired, e.g. to achieve recombination or other mutagenic procedures along with end-selection.

In accord with this invention, it is also appreciated that methods (e.g. mutagenesis methods) can be used to remove unwanted internal restriction sites. It is also appreciated that a partial digestion reaction (i.e. a digestion reaction that proceeds to partial completion) can be used to achieve digestion at a recognition site in a terminal region while sparing a susceptible restriction site that occurs internally in a polynucleotide and that is recognized by the same enzyme. In one aspect, partial digest are useful because it is appreciated that certain enzymes show preferential cleavage of the same recognition sequence depending on the location and environment in which the recognition sequence occurs. For example, it is appreciated that, while lambda DNA has 5 *EcoR* I sites, cleavage of the site nearest to the right terminus has been reported to occur 10 times faster than the sites in the middle of the molecule. Also, for example, it has been reported that, while *Sac* II has four sites on lambda DNA, the three clustered centrally in lambda are cleaved 50 times faster than the remaining site near the terminus (at nucleotide 40,386). Summarily, site preferences have been reported for various enzymes by many investigators (e.g., **Thomas and Davis, 1975; Forsblum et al, 1976; Nath and Azzolina, 1981; Brown and Smith, 1977; Gingeras and Brooks, 1983; Krüger et al, 1988; Conrad and Topal, 1989; Oller et al, 1991; Topal, 1991; and Pein, 1991;** to name but a few). It is appreciated that any empirical observations as well as any mechanistic understandings of site preferences by any serviceable polynucleotide-acting enzymes, whether currently available or to be procured in the future, may be serviceable in end-selection according to this invention.

It is also appreciated that protection methods can be used to selectively protect specified restriction sites (e.g. internal sites) against unwanted digestion by enzymes that would otherwise cut a working polypeptide in response to the presence of those sites; and that such protection methods include modifications such as methylations and base
5 substitutions (e.g. U instead of T) that inhibit an unwanted enzyme activity. It is appreciated that there are limited numbers of available restriction enzymes that are rare enough (e.g. having very long recognition sequences) to create large (e.g. megabase-long) restriction fragments, and that protection approaches (e.g. by methylation) are serviceable for increasing the rarity of enzyme cleavage sites. The use of *M.Fnu* II (mCGCG) to
10 increase the apparent rarity of *Not* I approximately twofold is but one example among many (Qiang et al, 1990; Nelson et al, 1984; Maxam and Gilbert, 1980; Raleigh and Wilson, 1986).

According to a preferred aspect of this invention, it is provided that, in general,
15 the use of rare restriction sites is preferred. It is appreciated that, in general, the frequency of occurrence of a restriction site is determined by the number of nucleotides contained therein, as well as by the ambiguity of the base requirements contained therein. Thus, in a non-limiting exemplification, it is appreciated that, in general, a restriction site composed of, for example, 8 specific nucleotides (e.g. the *Not* I site or GC/GGCCGC,
20 with an estimated relative occurrence of 1 in 4^8 , i.e. 1 in 65,536, random 8-mers) is relatively more infrequent than one composed of, for example, 6 nucleotides (e.g. the *Sma* I site or CCC/GGG, having an estimated relative occurrence of 1 in 4^6 , i.e. 1 in 4,096, random 6-mers), which in turn is relatively more infrequent than one composed of, for example, 4 nucleotides (e.g. the *Msp* I site or C/CGG, having an estimated relative
25 occurrence of 1 in 4^4 , i.e. 1 in 256, random 4-mers). Moreover, in another non-limiting exemplification, it is appreciated that, in general, a restriction site having no ambiguous (but only specific) base requirements (e.g. the *Fin* I site or GTCCC, having an estimated relative occurrence of 1 in 4^5 , i.e. 1 in 1024, random 5-mers) is relatively more infrequent than one having an ambiguous W (where W = A or T) base requirement (e.g. the *Ava* II

site or G/GWCC, having an estimated relative occurrence of 1 in $4 \times 4 \times 2 \times 4 \times 4$ - i.e. 1 in 512 - random 5-mers), which in turn is relatively more infrequent than one having an ambiguous N (where N = A or C or G or T) base requirement (e.g. the *Asu* I site or G/GNCC, having an estimated relative occurrence of 1 in $4 \times 4 \times 1 \times 4 \times 4$, i.e. 1 in 256 - random 5-mers). These relative occurrences are considered general estimates for actual polynucleotides, because it is appreciated that specific nucleotide bases (not to mention specific nucleotide sequences) occur with dissimilar frequencies in specific polynucleotides, in specific species of organisms, and in specific groupings of organisms. For example, it is appreciated that the % G+C contents of different species of organisms are often very different and wide ranging.

The use of relatively more infrequent restriction sites as a selection marker include - in a non-limiting fashion - preferably those sites composed at least a 4 nucleotide sequence, more preferably those composed at least a 5 nucleotide sequence, more preferably still those composed at least a 6 nucleotide sequence (e.g. the *Bam*H I site or G/GATCC, the *Bgl* II site or A/GATCT, the *Pst* I site or CTGCA/G, and the *Xba* I site or T/CTAGA), more preferably still those composed at least a 7 nucleotide sequence, more preferably still those composed of an 8 nucleotide sequence (e.g. the *Asc* I site or GG/CGCGCC, the *Not* I site or GC/GGCCGC, the *Pac* I site or TTAAT/TAA, the *Pme* I site or GTTT/AAAC, the *Srf* I site or GCCC/GGGC, the *Sse*838 I site or CCTGCA/GG, and the *Swa* I site or ATTT/AAAT), more preferably still those composed of a 9 nucleotide sequence, and even more preferably still those composed of at least a 10 nucleotide sequence (e.g. the *Bsp*G I site or CG/CGCTGGAC). It is further appreciated that some restriction sites (e.g. for class IIS enzymes) are comprised of a portion of relatively high specificity (i.e. a portion containing a principal determinant of the frequency of occurrence of the restriction site) and a portion of relatively low specificity; and that a site of cleavage may or may not be contained within a portion of relatively low specificity. For example, in the *Eco*57 I site or CTGAAG(16/14), there is a

portion of relatively high specificity (i.e. the CTGAAG portion) and a portion of relatively low specificity (i.e. the N16 sequence) that contains a site of cleavage.

In another preferred embodiment of this invention, a serviceable end-selection
5 marker is a terminal sequence that is recognized by a polynucleotide-acting enzyme that recognizes a specific polynucleotide sequence. In a preferred aspect of this invention, serviceable polynucleotide-acting enzymes also include other enzymes in addition to classic type II restriction enzymes. According to this preferred aspect of this invention, serviceable polynucleotide-acting enzymes also include gyrases, helicases, recombinases,
10 relaxases, and any enzymes related thereto.

Among preferred examples are topoisomerases (which have been categorized by some as a subset of the gyrases) and any other enzymes that have polynucleotide-cleaving activity (including preferably polynucleotide-nicking activity) &/or
15 polynucleotide-ligating activity. Among preferred topoisomerase enzymes are topoisomerase I enzymes, which is available from many commercial sources (Epicentre Technologies, Madison, WI; Invitrogen, Carlsbad, CA; Life Technologies, Gathesburg, MD) and conceivably even more private sources. It is appreciated that similar enzymes may be developed in the future that are serviceable for end-selection as provided herein.
20 A particularly preferred topoisomerase I enzyme is a topoisomerase I enzyme of vaccinia virus origin, that has a specific recognition sequence (e.g. 5'...AAGGG...3') and has both polynucleotide-nicking activity and polynucleotide-ligating activity. Due to the specific nicking-activity of this enzyme (cleavage of one strand), internal recognition sites are not prone to polynucleotide destruction resulting from the nicking activity (but
25 rather remain annealed) at a temperature that causes denaturation of a terminal site that has been nicked. Thus for use in end-selection, it is preferable that a nicking site for topoisomerase-based end-selection be no more than 100 nucleotides from a terminus, more preferably no more than 50 nucleotides from a terminus, more preferably still no more than 25 nucleotides from a terminus, even more preferably still no more than 20

nucleotides from a terminus, even more preferably still no more than 15 nucleotides from a terminus, even more preferably still no more than 10 nucleotides from a terminus, even more preferably still no more than 8 nucleotides from a terminus, even more preferably still no more than 6 nucleotides from a terminus, and even more preferably still no more than 4 nucleotides from a terminus.

In a particularly preferred exemplification that is non-limiting yet clearly illustrative, it is appreciated that when a nicking site for topoisomerase-based end-selection is 4 nucleotides from a terminus, nicking produces a single stranded oligo of 4 bases (in a terminal region) that can be denatured from its complementary strand in an end-selectable polynucleotide; this provides a sticky end (comprised of 4 bases) in a polynucleotide that is serviceable for an ensuing ligation reaction. To accomplish ligation to a cloning vector (preferably an expression vector), compatible sticky ends can be generated in a cloning vector by any means including by restriction enzyme-based means. The terminal nucleotides (comprised of 4 terminal bases in this specific example) in an end-selectable polynucleotide terminus are thus wisely chosen to provide compatibility with a sticky end generated in a cloning vector to which the polynucleotide is to be ligated.

On the other hand, internal nicking of an end-selectable polynucleotide, e.g. 500 bases from a terminus, produces a single stranded oligo of 500 bases that is not easily denatured from its complementary strand, but rather is serviceable for repair (e.g. by the same topoisomerase enzyme that produced the nick).

This invention thus provides a method - e.g. that is vaccinia topoisomerase-based &/or type II (or IIS) restriction endonuclease-based &/or type III restriction endonuclease-based &/or nicking enzyme-based (e.g. using N. *Bsr*NB I) - for producing a sticky end in a working polynucleotide, which end is ligation compatible, and which end can be comprised of at least a 1 base overhang. Preferably such a sticky end is

comprised of at least a 2-base overhang, more preferably such a sticky end is comprised of at least a 3-base overhang, more preferably still such a sticky end is comprised of at least a 4-base overhang, even more preferably still such a sticky end is comprised of at least a 5-base overhang, even more preferably still such a sticky end is comprised of at least a 6-base overhang. Such a sticky end may also be comprised of at least a 7-base overhang, or at least an 8-base overhang, or at least a 9-base overhang, or at least a 10-base overhang, or at least 15-base overhang, or at least a 20-base overhang, or at least a 25-base overhang, or at least a 30-base overhang. These overhangs can be comprised of any bases, including A, C, G, or T.

10

It is appreciated that sticky end overhangs introduced using topoisomerase or a nicking enzyme (e.g. using *N. Bst* I) can be designed to be unique in a ligation environment, so as to prevent unwanted fragment reassemblies, such as self-dimerizations and other unwanted concatamerizations.

15

According to one aspect of this invention, a plurality of sequences (which may but do not necessarily overlap) can be introduced into a terminal region of an end-selectable polynucleotide by the use of an oligo in a polymerase-based reaction. In a relevant, but by no means limiting example, such an oligo can be used to provide a preferred 5' terminal region that is serviceable for topoisomerase I-based end-selection, which oligo is comprised of: a 1-10 base sequence that is convertible into a sticky end (preferably by a vaccinia topoisomerase I), a ribosome binding site (i.e. and "RBS", that is preferably serviceable for expression cloning), and optional linker sequence followed by an ATG start site and a template-specific sequence of 0-100 bases (to facilitate annealment to the template in the a polymerase-based reaction). Thus, according to this example, a serviceable oligo (which may be termed a forward primer) can have the sequence:

5'[terminal sequence = (N)₁₋₁₀][topoisomerase I site & RBS = AAGGGAGGAG][linker = (N)₁₋₁₀₀][start codon and template-specific sequence = ATG(N)₀₋₁₀₀3'.

25

Analogously, in a relevant, but by no means limiting example, an oligo can be used to provide a preferred 3' terminal region that is serviceable for topoisomerase I-based end-selection, which oligo is comprised of: a 1-10 base sequence that is convertible into a sticky end (preferably by a vaccinia topoisomerase I), and optional linker sequence
5 followed by a template-specific sequence of 0-100 bases (to facilitate annealment to the template in the a polymerase-based reaction). Thus, according to this example, a serviceable oligo (which may be termed a reverse primer) can have the sequence:
5'[terminal sequence = (N)₁₋₁₀][topoisomerase I site = AAGGG][linker = (N)₁₋₁₀₀][template-specific sequence = (N)₀₋₁₀₀]3'.

10

It is appreciated that, end-selection can be used to distinguish and separate parental template molecules (e.g. to be subjected to mutagenesis) from progeny molecules (e.g. generated by mutagenesis). For example, a first set of primers, lacking in a topoisomerase I recognition site, can be used to modify the terminal regions of the parental molecules (e.g. in
15 polymerase-based amplification). A different second set of primers (e.g. having a topoisomerase I recognition site) can then be used to generate mutated progeny molecules (e.g. using any polynucleotide chimerization method, such as interrupted synthesis, template-switching polymerase-based amplification, or interrupted synthesis; or using saturation mutagenesis; or using any other method for introducing a topoisomerase I recognition site into
20 a mutagenized progeny molecule as disclosed herein) from the amplified template molecules. The use of topoisomerase I-based end-selection can then facilitate, not only discernment, but selective topoisomerase I-based ligation of the desired progeny molecules.

Annealment of a second set of primers to thusly amplified parental molecules can be
25 facilitated by including sequences in a first set of primers (i.e. primers used for amplifying a set parental molecules) that are similar to a topoisomerase I recognition site, yet different enough to prevent functional topoisomerase I enzyme recognition. For example, sequences that diverge from the AAGGG site by anywhere from 1 base to all 5 bases can be incorporated into a first set of primers (to be used for amplifying the parental templates prior to subjection

to mutagenesis). In a specific, but non-limiting aspect, it is thus provided that a parental molecule can be amplified using the following exemplary – but by no means limiting – set of forward and reverse primers:

- 5 Forward Primer: 5' CTAGAAGAGAGGAGAAAACCATG(N)₁₀₋₁₀₀ 3', and
Reverse Primer: 5' GATCAAAGGCGCGCCTGCAGG(N)₁₀₋₁₀₀ 3'

According to this specific example of a first set of primers, (N)₁₀₋₁₀₀ represents preferably a 10 to 100 nucleotide-long template-specific sequence, more preferably a 10 to 50
10 nucleotide-long template-specific sequence, more preferably still a 10 to 30 nucleotide-long template-specific sequence, and even more preferably still a 15 to 25 nucleotide-long template-specific sequence.

According to a specific, but non-limiting aspect, it is thus provided that, after this
15 amplification (using a disclosed first set of primers lacking in a true topoisomerase I recognition site), amplified parental molecules can then be subjected to mutagenesis using one or more sets of forward and reverse primers that do have a true topoisomerase I recognition site. In a specific, but non-limiting aspect, it is thus provided that a parental molecule can be used as templates for the generation of a mutagenized progeny molecule using the following
20 exemplary – but by no means limiting – second set of forward and reverse primers:

Forward Primer: 5' CTAGAAGGGAGGAGAAAACCATG 3'

Reverse Primer: 5' GATCAAAGGCGCGCCTGCAGG 3' (contains *Asc* I
recognition sequence)

25

It is appreciated that any number of different primers sets not specifically mentioned can be used as first, second, or subsequent sets of primers for end-selection consistent with this invention. Notice that type II restriction enzyme sites can be incorporated (e.g. an *Asc* I site in the above example). It is provided that, in addition to the other sequences mentioned,

the experimentalist can incorporate one or more N,N,G/T triplets into a serviceable primer in order to subject a working polynucleotide to saturation mutagenesis. Summarily, use of a second and/or subsequent set of primers can achieve dual goals of introducing a topoisomerase I site and of generating mutations in a progeny polynucleotide.

5

Thus, according to one use provided, a serviceable end-selection marker is an enzyme recognition site that allows an enzyme to cleave (including nick) a polynucleotide at a specified site, to produce a ligation-compatible end upon denaturation of a generated single stranded oligo. Ligation of the produced polynucleotide end can then be accomplished by the same enzyme (e.g. in the case of vaccinia virus topoisomerase I), or alternatively with the use of a different enzyme. According to one aspect of this invention, any serviceable end-selection markers, whether like (e.g. two vaccinia virus topoisomerase I recognition sites) or unlike (e.g. a class II restriction enzyme recognition site and a vaccinia virus topoisomerase I recognition site) can be used in combination to select a polynucleotide. Each selectable polynucleotide can thus have one or more end-selection markers, and they can be like or unlike end-selection markers. In a particular aspect, a plurality of end-selection markers can be located on one end of a polynucleotide and can have overlapping sequences with each other.

It is important to emphasize that any number of enzymes, whether currently in existence or to be developed, can be serviceable in end-selection according to this invention. For example, in a particular aspect of this invention, a nicking enzyme (e.g. N. *Bst*NB I, which cleaves only one strand at 5'...GAGTCNNNN/N...3') can be used in conjunction with a source of polynucleotide-ligating activity in order to achieve end-selection. According to this embodiment, a recognition site for N. *Bst*NB I – instead of a recognition site for topoisomerase I – should be incorporated into an end-selectable polynucleotide (whether end-selection is used for selection of a mutagenized progeny molecule or whether end-selection is used apart from any mutagenesis procedure).

It is appreciated that the instantly disclosed end-selection approach using topoisomerase-based nicking and ligation has several advantages over previously available selection methods. In sum, this approach allows one to achieve direction cloning (including expression cloning). Specifically, this approach can be used for the achievement of:

5 direct ligation (i.e. without subjection to a classic restriction-purification-ligation reaction, that is susceptible to a multitude of potential problems from an initial restriction reaction to a ligation reaction dependent on the use of T4 DNA ligase); separation of progeny molecules from original template molecules (e.g. original template molecules lack topoisomerase I sites that not introduced until after mutagenesis), obviation of the

10 need for size separation steps (e.g. by gel chromatography or by other electrophoretic means or by the use of size-exclusion membranes), preservation of internal sequences (even when topoisomerase I sites are present), obviation of concerns about unsuccessful ligation reactions (e.g. dependent on the use of T4 DNA ligase, particularly in the presence of unwanted residual restriction enzyme activity), and facilitated expression

15 cloning (including obviation of frame shift concerns). Concerns about unwanted restriction enzyme-based cleavages – especially at internal restriction sites (or even at often unpredictable sites of unwanted star activity) in a working polynucleotide – that are potential sites of destruction of a working polynucleotide can also be obviated by the instantly disclosed end-selection approach using topoisomerase-based nicking and

20 ligation.

2.11.3. ADDITIONAL SCREENING METHODS

25 Peptide Display Methods

The present method can be used to shuffle, by *in vitro* and/or *in vivo* recombination by any of the disclosed methods, and in any combination, polynucleotide sequences selected by peptide display methods, wherein an associated polynucleotide encodes a displayed peptide which is screened for a phenotype (e.g., for affinity for a

30 predetermined receptor (ligand).

An increasingly important aspect of bio-pharmaceutical drug development and molecular biology is the identification of peptide structures, including the primary amino acid sequences, of peptides or peptidomimetics that interact with biological
5 macromolecules. one method of identifying peptides that possess a desired structure or functional property, such as binding to a predetermined biological macromolecule (e.g., a receptor), involves the screening of a large library or peptides for individual library members which possess the desired structure or functional property conferred by the amino acid sequence of the peptide.

10

In addition to direct chemical synthesis methods for generating peptide libraries, several recombinant DNA methods also have been reported. One type involves the display of a peptide sequence, antibody, or other protein on the surface of a bacteriophage particle or cell. Generally, in these methods each bacteriophage particle or cell serves as
15 an individual library member displaying a single species of displayed peptide in addition to the natural bacteriophage or cell protein sequences. Each bacteriophage or cell contains the nucleotide sequence information encoding the particular displayed peptide sequence; thus, the displayed peptide sequence can be ascertained by nucleotide sequence determination of an isolated library member.

20

A well-known peptide display method involves the presentation of a peptide sequence on the surface of a filamentous bacteriophage, typically as a fusion with a bacteriophage coat protein. The bacteriophage library can be incubated with an immobilized, predetermined macromolecule or small molecule (e.g., a receptor) so that
25 bacteriophage particles which present a peptide sequence that binds to the immobilized macromolecule can be differentially partitioned from those that do not present peptide sequences that bind to the predetermined macromolecule. The bacteriophage particles (i.e., library members) which are bound to the immobilized macromolecule are then recovered and replicated to amplify the selected bacteriophage sub-population for a

subsequent round of affinity enrichment and phage replication. After several rounds of affinity enrichment and phage replication, the bacteriophage library members that are thus selected are isolated and the nucleotide sequence encoding the displayed peptide sequence is determined, thereby identifying the sequence(s) of peptides that bind to the predetermined macromolecule (e.g., receptor). Such methods are further described in PCT patent publications WO 91/17271, WO 91/18980, WO 91/19818 and WO 93/08278.

The latter PCT publication describes a recombinant DNA method for the display of peptide ligands that involves the production of a library of fusion proteins with each fusion protein composed of a first polypeptide portion, typically comprising a variable sequence, that is available for potential binding to a predetermined macromolecule, and a second polypeptide portion that binds to DNA, such as the DNA vector encoding the individual fusion protein. When transformed host cells are cultured under conditions that allow for expression of the fusion protein, the fusion protein binds to the DNA vector encoding it. Upon lysis of the host cell, the fusion protein/vector DNA complexes can be screened against a predetermined macromolecule in much the same way as bacteriophage particles are screened in the phage-based display system, with the replication and sequencing of the DNA vectors in the selected fusion protein/vector DNA complexes serving as the basis for identification of the selected library peptide sequence(s).

Other systems for generating libraries of peptides and like polymers have aspects of both the recombinant and *in vitro* chemical synthesis methods. In these hybrid methods, cell-free enzymatic machinery is employed to accomplish the *in vitro* synthesis of the library members (i.e., peptides or polynucleotides). In one type of method, RNA molecules with the ability to bind a predetermined protein or a predetermined dye molecule were selected by alternate rounds of selection and PCR amplification (Tuerk and Gold, 1990; Ellington and Szostak, 1990). A similar technique was used to identify DNA sequences which bind a predetermined human transcription factor

(Thiesen and Bach, 1990; Beaudry and Joyce, 1992; PCT patent publications WO 92/05258 and WO 92/14843). In a similar fashion, the technique of *in vitro* translation has been used to synthesize proteins of interest and has been proposed as a method for generating large libraries of peptides. These methods which rely upon *in vitro* translation, generally comprising stabilized polysome complexes, are described further in PCT patent publications WO 88/08453, WO 90/05785, WO 90/07003, WO 91/02076, WO 91/05058, and WO 92/02536. Applicants have described methods in which library members comprise a fusion protein having a first polypeptide portion with DNA binding activity and a second polypeptide portion having the library member unique peptide sequence; such methods are suitable for use in cell-free *in vitro* selection formats, among others.

The displayed peptide sequences can be of varying lengths, typically from 3-5000 amino acids long or longer, frequently from 5-100 amino acids long, and often from about 8-15 amino acids long. A library can comprise library members having varying lengths of displayed peptide sequence, or may comprise library members having a fixed length of displayed peptide sequence. Portions or all of the displayed peptide sequence(s) can be random, pseudorandom, defined set kernel, fixed, or the like. The present display methods include methods for *in vitro* and *in vivo* display of single-chain antibodies, such as nascent scFv on polysomes or scFv displayed on phage, which enable large-scale screening of scFv libraries having broad diversity of variable region sequences and binding specificities.

The present invention also provides random, pseudorandom, and defined sequence framework peptide libraries and methods for generating and screening those libraries to identify useful compounds (e.g., peptides, including single-chain antibodies) that bind to receptor molecules or epitopes of interest or gene products that modify peptides or RNA in a desired fashion. The random, pseudorandom, and defined sequence framework peptides are produced from libraries of peptide library members that comprise

displayed peptides or displayed single-chain antibodies attached to a polynucleotide template from which the displayed peptide was synthesized. The mode of attachment may vary according to the specific embodiment of the invention selected, and can include encapsulation in a phage particle or incorporation in a cell.

5

A method of affinity enrichment allows a very large library of peptides and single-chain antibodies to be screened and the polynucleotide sequence encoding the desired peptide(s) or single-chain antibodies to be selected. The polynucleotide can then be isolated and shuffled to recombine combinatorially the amino acid sequence of the
10 selected peptide(s) (or predetermined portions thereof) or single-chain antibodies (or just VHI, VLI or CDR portions thereof). Using these methods, one can identify a peptide or single-chain antibody as having a desired binding affinity for a molecule and can exploit the process of shuffling to converge rapidly to a desired high-affinity peptide or scfv. The peptide or antibody can then be synthesized in bulk by conventional means for any
15 suitable use (e.g., as a therapeutic or diagnostic agent).

A significant advantage of the present invention is that no prior information regarding an expected ligand structure is required to isolate peptide ligands or antibodies of interest. The peptide identified can have biological activity, which is meant to include
20 at least specific binding affinity for a selected receptor molecule and, in some instances, will further include the ability to block the binding of other compounds, to stimulate or inhibit metabolic pathways, to act as a signal or messenger, to stimulate or inhibit cellular activity, and the like.

25 The present invention also provides a method for shuffling a pool of polynucleotide sequences selected by affinity screening a library of polysomes displaying nascent peptides (including single-chain antibodies) for library members which bind to a predetermined receptor (e.g., a mammalian proteinaceous receptor such as, for example, a peptidergic hormone receptor, a cell surface receptor, an intracellular protein which binds

to other protein(s) to form intracellular protein complexes such as hetero-dimers and the like) or epitope (e.g., an immobilized protein, glycoprotein, oligosaccharide, and the like).

- 5 Polynucleotide sequences selected in a first selection round (typically by affinity selection for binding to a receptor (e.g., a ligand)) by any of these methods are pooled and the pool(s) is/are shuffled by *in vitro* and/or *in vivo* recombination to produce a shuffled pool comprising a population of recombined selected polynucleotide sequences. The recombined selected polynucleotide sequences are subjected to at least one
- 10 subsequent selection round. The polynucleotide sequences selected in the subsequent selection round(s) can be used directly, sequenced, and/or subjected to one or more additional rounds of shuffling and subsequent selection. Selected sequences can also be back-crossed with polynucleotide sequences encoding neutral sequences (i.e., having insubstantial functional effect on binding), such as for example by back-crossing with a
- 15 wild-type or naturally-occurring sequence substantially identical to a selected sequence to produce native-like functional peptides, which may be less immunogenic. Generally, during back-crossing subsequent selection is applied to retain the property of binding to the predetermined receptor (ligand).

- 20 Prior to or concomitant with the shuffling of selected sequences, the sequences can be mutagenized. In one embodiment, selected library members are cloned in a prokaryotic vector (e.g., plasmid, phagemid, or bacteriophage) wherein a collection of individual colonies (or plaques) representing discrete library members are produced. Individual selected library members can then be manipulated (e.g., by site-directed
- 25 mutagenesis, cassette mutagenesis, chemical mutagenesis, PCR mutagenesis, and the like) to generate a collection of library members representing a kernel of sequence diversity based on the sequence of the selected library member. The sequence of an individual selected library member or pool can be manipulated to incorporate random mutation, pseudorandom mutation, defined kernel mutation (i.e., comprising variant and

invariant residue positions and/or comprising variant residue positions which can comprise a residue selected from a defined subset of amino acid residues), codon-based mutation, and the like, either segmentally or over the entire length of the individual selected library member sequence. The mutagenized selected library members are then
5 shuffled by *in vitro* and/or *in vivo* recombinatorial shuffling as disclosed herein.

The invention also provides peptide libraries comprising a plurality of individual library members of the invention, wherein (1) each individual library member of said plurality comprises a sequence produced by shuffling of a pool of selected sequences, and
10 (2) each individual library member comprises a variable peptide segment sequence or single-chain antibody segment sequence which is distinct from the variable peptide segment sequences or single-chain antibody sequences of other individual library members in said plurality (although some library members may be present in more than one copy per library due to uneven amplification, stochastic probability, or the like).

15

The invention also provides a product-by-process, wherein selected polynucleotide sequences having (or encoding a peptide having) a predetermined binding specificity are formed by the process of: (1) screening a displayed peptide or displayed single-chain antibody library against a predetermined receptor (e.g., ligand) or epitope
20 (e.g., antigen macromolecule) and identifying and/or enriching library members which bind to the predetermined receptor or epitope to produce a pool of selected library members, (2) shuffling by recombination the selected library members (or amplified or cloned copies thereof) which binds the predetermined epitope and has been thereby isolated and/or enriched from the library to generate a shuffled library, and (3) screening
25 the shuffled library against the predetermined receptor (e.g., ligand) or epitope (e.g., antigen macromolecule) and identifying and/or enriching shuffled library members which bind to the predetermined receptor or epitope to produce a pool of selected shuffled library members.

Antibody Display and Screening Methods

The present method can be used to shuffle, by *in vitro* and/or *in vivo* recombination by any of the disclosed methods, and in any combination, polynucleotide sequences selected by antibody display methods, wherein an associated polynucleotide encodes a displayed antibody which is screened for a phenotype (e.g., for affinity for binding a predetermined antigen (ligand)).

Various molecular genetic approaches have been devised to capture the vast immunological repertoire represented by the extremely large number of distinct variable regions which can be present in immunoglobulin chains. The naturally-occurring germ line immunoglobulin heavy chain locus is composed of separate tandem arrays of variable segment genes located upstream of a tandem array of diversity segment genes, which are themselves located upstream of a tandem array of joining (i) region genes, which are located upstream of the constant region genes. During B lymphocyte development, V-D-J rearrangement occurs wherein a heavy chain variable region gene (VH) is formed by rearrangement to form a fused D segment followed by rearrangement with a V segment to form a V-D-J joined product gene which, if productively rearranged, encodes a functional variable region (VH) of a heavy chain. Similarly, light chain loci rearrange one of several V segments with one of several J segments to form a gene encoding the variable region (VL) of a light chain.

The vast repertoire of variable regions possible in immunoglobulins derives in part from the numerous combinatorial possibilities of joining V and i segments (and, in the case of heavy chain loci, D segments) during rearrangement in B cell development. Additional sequence diversity in the heavy chain variable regions arises from non-uniform rearrangements of the D segments during V-D-J joining and from N region addition. Further, antigen-selection of specific B cell clones selects for higher affinity variants having non-germline mutations in one or both of the heavy and light chain

variable regions; a phenomenon referred to as "affinity maturation" or "affinity sharpening". Typically, these "affinity sharpening" mutations cluster in specific areas of the variable region, most commonly in the complementarity-determining regions (CDRs).

- 5 In order to overcome many of the limitations in producing and identifying high-affinity immunoglobulins through antigen-stimulated β cell development (i.e., immunization), various prokaryotic expression systems have been developed that can be manipulated to produce combinatorial antibody libraries which may be screened for high-affinity antibodies to specific antigens. Recent advances in the expression of
- 10 antibodies in *Escherichia coli* and bacteriophage systems (see "alternative peptide display methods", *infra*) have raised the possibility that virtually any specificity can be obtained by either cloning antibody genes from characterized hybridomas or by de novo selection using antibody gene libraries (e.g., from Ig cDNA).
- 15 Combinatorial libraries of antibodies have been generated in bacteriophage lambda expression systems which may be screened as bacteriophage plaques or as colonies of lysogens (Huse et al, 1989; Caton and Koprowski, 1990; Mullinax et al, 1990; Persson et al, 1991). Various embodiments of bacteriophage antibody display libraries and lambda phage expression libraries have been described (Kang et al, 1991;
- 20 Clackson et al, 1991; McCafferty et al, 1990; Burton et al, 1991; Hoogenboom et al, 1991; Chang et al, 1991; Breitling et al, 1991; Marks et al, 1991, p. 581; Barbas et al, 1992; Hawkins and Winter, 1992; Marks et al, 1992, p. 779; Marks et al, 1992, p. 16007; and Lowman et al, 1991; Lerner et al, 1992; all incorporated herein by reference). Typically, a bacteriophage antibody display library is screened with a receptor
- 25 (e.g., polypeptide, carbohydrate, glycoprotein, nucleic acid) that is immobilized (e.g., by covalent linkage to a chromatography resin to enrich for reactive phage by affinity chromatography) and/or labeled (e.g., to screen plaque or colony lifts).

One particularly advantageous approach has been the use of so-called single-chain fragment variable (scfv) libraries (Marks et al, 1992, p. 779; Winter and Milstein, 1991; Clackson et al, 1991; Marks et al, 1991, p. 581; Chaudhary et al, 1990; Chiswell et al, 1992; McCafferty et al, 1990; and Huston et al, 1988). Various embodiments of scfv
5 libraries displayed on bacteriophage coat proteins have been described.

Beginning in 1988, single-chain analogues of Fv fragments and their fusion proteins have been reliably generated by antibody engineering methods. The first step generally involves obtaining the genes encoding VH and VL domains with desired
10 binding properties; these V genes may be isolated from a specific hybridoma cell line, selected from a combinatorial V-gene library, or made by V gene synthesis. The single-chain Fv is formed by connecting the component V genes with an oligonucleotide that encodes an appropriately designed linker peptide, such as (Gly-Gly-Gly-Gly-Ser)₃ or equivalent linker peptide(s). The linker bridges the C-terminus of the first V region and
15 N-terminus of the second, ordered as either VH-linker-VL or VL-linker-VH. In principle, the scfv binding site can faithfully replicate both the affinity and specificity of its parent antibody combining site.

Thus, scfv fragments are comprised of VH and VL domains linked into a single
20 polypeptide chain by a flexible linker peptide. After the scfv genes are assembled, they are cloned into a phagemid and expressed at the tip of the M13 phage (or similar filamentous bacteriophage) as fusion proteins with the bacteriophage PIII (gene 3) coat protein. Enriching for phage expressing an antibody of interest is accomplished by panning the recombinant phage displaying a population scfv for binding to a
25 predetermined epitope (e.g., target antigen, receptor).

The linked polynucleotide of a library member provides the basis for replication of the library member after a screening or selection procedure, and also provides the basis for the determination, by nucleotide sequencing, of the identity of the displayed peptide

sequence or VH and VL amino acid sequence. The displayed peptide (s) or single-chain antibody (e. g., scfv) and/or its VH and VL domains or their CDRs can be cloned and expressed in a suitable expression system. Often polynucleotides encoding the isolated VH and VL domains will be ligated to polynucleotides encoding constant regions (CH and CL) to form polynucleotides encoding complete antibodies (e.g., chimeric or fully-human), antibody fragments, and the like. Often polynucleotides encoding the isolated CDRs will be grafted into polynucleotides encoding a suitable variable region framework (and optionally constant regions) to form polynucleotides encoding complete antibodies (e.g., humanized or fully-human), antibody fragments, and the like.

10 Antibodies can be used to isolate preparative quantities of the antigen by immunoaffinity chromatography. Various other uses of such antibodies are to diagnose and/or stage disease (e.g., neoplasia) and for therapeutic application to treat disease, such as for example: neoplasia, autoimmune disease, AIDS, cardiovascular disease, infections, and the like.

15

Various methods have been reported for increasing the combinatorial diversity of a scfv library to broaden the repertoire of binding species (idiotype spectrum) The use of PCR has permitted the variable regions to be rapidly cloned either from a specific hybridoma source or as a gene library from non-immunized cells, affording combinatorial diversity in the assortment of VH and VL cassettes which can be combined.

20 Furthermore, the VH and VL cassettes can themselves be diversified, such as by random, pseudorandom, or directed mutagenesis. Typically, VH and VL cassettes are diversified in or near the complementarity-determining regions (CDRS), often the third CDR, CDR3. Enzymatic inverse PCR mutagenesis has been shown to be a simple and reliable method

25 for constructing relatively large libraries of scfv site-directed hybrids (Stemmer et al, 1993), as has error-prone PCR and chemical mutagenesis (Deng et al, 1994). Riechmann (Riechmann et al, 1993) showed semi-rational design of an antibody scfv fragment using site-directed randomization by degenerate oligonucleotide PCR and subsequent phage display of the resultant scfv hybrids. Barbas (Barbas et al, 1992) attempted to circumvent

the problem of limited repertoire sizes resulting from using biased variable region sequences by randomizing the sequence in a synthetic CDR region of a human tetanus toxoid-binding Fab.

5 CDR randomization has the potential to create approximately 1×10^{20} CDRs for the heavy chain CDR3 alone, and a roughly similar number of variants of the heavy chain CDR1 and CDR2, and light chain CDR1-3 variants. Taken individually or together, the combination possibilities of CDR randomization of heavy and/or light chains requires generating a prohibitive number of bacteriophage clones to produce a clone library
10 representing all possible combinations, the vast majority of which will be non-binding. Generation of such large numbers of primary transformants is not feasible with current transformation technology and bacteriophage display systems. For example, Barbas (Barbas et al, 1992) only generated 5×10^7 transformants, which represents only a tiny fraction of the potential diversity of a library of thoroughly randomized CDRS.

15 Despite these substantial limitations, bacteriophage display of scfv have already yielded a variety of useful antibodies and antibody fusion proteins. A bispecific single chain antibody has been shown to mediate efficient tumor cell lysis (Gruber et al, 1994). Intracellular expression of an anti-Rev scfv has been shown to inhibit HIV-1 virus
20 replication *in vitro* (Duan et al, 1994), and intracellular expression of an anti-p21^{lar}, scfv has been shown to inhibit meiotic maturation of Xenopus oocytes (Biocca et al, 1993). Recombinant scfv which can be used to diagnose HIV infection have also been reported, demonstrating the diagnostic utility of scfv (Lilley et al, 1994). Fusion proteins wherein an scFv is linked to a second polypeptide, such as a toxin or fibrinolytic activator protein,
25 have also been reported (Holvast et al, 1992; Nicholls et al, 1993).

If it were possible to generate scfv libraries having broader antibody diversity and overcoming many of the limitations of conventional CDR mutagenesis and randomization methods which can cover only a very tiny fraction of the potential

sequence combinations, the number and quality of scfv antibodies suitable for therapeutic and diagnostic use could be vastly improved. To address this, the *in vitro* and *in vivo* shuffling methods of the invention are used to recombine CDRs which have been obtained (typically via PCR amplification or cloning) from nucleic acids obtained from
5 selected displayed antibodies. Such displayed antibodies can be displayed on cells, on bacteriophage particles, on polysomes, or any suitable antibody display system wherein the antibody is associated with its encoding nucleic acid(s). In a variation, the CDRs are initially obtained from mRNA (or cDNA) from antibody-producing cells (e.g., plasma cells/splenocytes from an immunized wild-type mouse, a human, or a transgenic mouse
10 capable of making a human antibody as in WO 92/03918, WO 93/12227, and WO 94/25585), including hybridomas derived therefrom.

Polynucleotide sequences selected in a first selection round (typically by affinity selection for displayed antibody binding to an antigen (e.g., a ligand) by any of these
15 methods are pooled and the pool(s) is/are shuffled by *in vitro* and/or *in vivo* recombination, especially shuffling of CDRs (typically shuffling heavy chain CDRs with other heavy chain CDRs and light chain CDRs with other light chain CDRs) to produce a shuffled pool comprising a population of recombined selected polynucleotide sequences. The recombined selected polynucleotide sequences are expressed in a selection format as
20 a displayed antibody and subjected to at least one subsequent selection round. The polynucleotide sequences selected in the subsequent selection round(s) can be used directly, sequenced, and/or subjected to one or more additional rounds of shuffling and subsequent selection until an antibody of the desired binding affinity is obtained. Selected sequences can also be back-crossed with polynucleotide sequences encoding
25 neutral antibody framework sequences (i.e., having insubstantial functional effect on antigen binding), such as for example by back-crossing with a human variable region framework to produce human-like sequence antibodies. Generally, during back-crossing subsequent selection is applied to retain the property of binding to the predetermined antigen.

Alternatively, or in combination with the noted variations, the valency of the target epitope may be varied to control the average binding affinity of selected scfv library members. The target epitope can be bound to a surface or substrate at varying
5 densities, such as by including a competitor epitope, by dilution, or by other method known to those in the art. A high density (valency) of predetermined epitope can be used to enrich for scfv library members which have relatively low affinity, whereas a low density (valency) can preferentially enrich for higher affinity scfv library members.

10 For generating diverse variable segments, a collection of synthetic oligonucleotides encoding random, pseudorandom, or a defined sequence kernel set of peptide sequences can be inserted by ligation into a predetermined site (e.g., a CDR). Similarly, the sequence diversity of one or more CDRs of the single-chain antibody cassette(s) can be expanded by mutating the CDR(s) with site-directed mutagenesis,
15 CDR-replacement, and the like. The resultant DNA molecules can be propagated in a host for cloning and amplification prior to shuffling, or can be used directly (i.e., may avoid loss of diversity which may occur upon propagation in a host cell) and the selected library members subsequently shuffled.

20 Displayed peptide/polynucleotide complexes (library members) which encode a variable segment peptide sequence of interest or a single-chain antibody of interest are selected from the library by an affinity enrichment technique. This is accomplished by means of a immobilized macromolecule or epitope specific for the peptide sequence of interest, such as a receptor, other macromolecule, or other epitope species. Repeating the
25 affinity selection procedure provides an enrichment of library members encoding the desired sequences, which may then be isolated for pooling and shuffling, for sequencing, and/or for further propagation and affinity enrichment.

The library members without the desired specificity are removed by washing. The degree and stringency of washing required will be determined for each peptide sequence or single-chain antibody of interest and the immobilized predetermined macromolecule or epitope. A certain degree of control can be exerted over the binding characteristics of the nascent peptide/DNA complexes recovered by adjusting the conditions of the binding incubation and the subsequent washing. The temperature, pH, ionic strength, divalent cations concentration, and the volume and duration of the washing will select for nascent peptide/DNA complexes within particular ranges of affinity for the immobilized macromolecule. Selection based on slow dissociation rate, which is usually predictive of high affinity, is often the most practical route. This may be done either by continued incubation in the presence of a saturating amount of free predetermined macromolecule, or by increasing the volume, number, and length of the washes. In each case, the rebinding of dissociated nascent peptide/DNA or peptide/RNA complex is prevented, and with increasing time, nascent peptide/DNA or peptide/RNA complexes of higher and higher affinity are recovered.

Additional modifications of the binding and washing procedures may be applied to find peptides with special characteristics. The affinities of some peptides are dependent on ionic strength or cation concentration. This is a useful characteristic for peptides that will be used in affinity purification of various proteins when gentle conditions for removing the protein from the peptides are required.

One variation involves the use of multiple binding targets (multiple epitope species, multiple receptor species), such that a scfv library can be simultaneously screened for a multiplicity of scfv which have different binding specificities. Given that the size of a scfv library often limits the diversity of potential scfv sequences, it is typically desirable to use scfv libraries of as large a size as possible. The time and economic considerations of generating a number of very large polysome scFv-display libraries can become prohibitive. To avoid this substantial problem, multiple

predetermined epitope species (receptor species) can be concomitantly screened in a single library, or sequential screening against a number of epitope species can be used. In one variation, multiple target epitope species, each encoded on a separate bead (or subset of beads), can be mixed and incubated with a polysome-display scfv library under
5 suitable binding conditions. The collection of beads, comprising multiple epitope species, can then be used to isolate, by affinity selection, scfv library members. Generally, subsequent affinity screening rounds can include the same mixture of beads, subsets thereof, or beads containing only one or two individual epitope species. This approach affords efficient screening, and is compatible with laboratory automation, batch
10 processing, and high throughput screening methods.

A variety of techniques can be used in the present invention to diversify a peptide library or single-chain antibody library, or to diversify, prior to or concomitant with shuffling, around variable segment peptides found in early rounds of panning to have
15 sufficient binding activity to the predetermined macromolecule or epitope. In one approach, the positive selected peptide/polynucleotide complexes (those identified in an early round of affinity enrichment) are sequenced to determine the identity of the active peptides. Oligonucleotides are then synthesized based on these active peptide sequences, employing a low level of all bases incorporated at each step to produce slight variations
20 of the primary oligonucleotide sequences. This mixture of (slightly) degenerate oligonucleotides is then cloned into the variable segment sequences at the appropriate locations. This method produces systematic, controlled variations of the starting peptide sequences, which can then be shuffled. It requires, however, that individual positive nascent peptide/polynucleotide complexes be sequenced before mutagenesis, and thus is
25 useful for expanding the diversity of small numbers of recovered complexes and selecting variants having higher binding affinity and/or higher binding specificity. In a variation, mutagenic PCR amplification of positive selected peptide/polynucleotide complexes (especially of the variable region sequences, the amplification products of which are shuffled *in vitro* and/or *in vivo* and one or more additional rounds of screening is done

prior to sequencing. The same general approach can be employed with single-chain antibodies in order to expand the diversity and enhance the binding affinity/specificity, typically by diversifying CDRs or adjacent framework regions prior to or concomitant with shuffling. If desired, shuffling reactions can be spiked with mutagenic oligonucleotides capable of *in vitro* recombination with the selected library members can be included. Thus, mixtures of synthetic oligonucleotides and PCR produced polynucleotides (synthesized by error-prone or high-fidelity methods) can be added to the *in vitro* shuffling mix and be incorporated into resulting shuffled library members (shufflants).

10

The present invention of shuffling enables the generation of a vast library of CDR-variant single-chain antibodies. One way to generate such antibodies is to insert synthetic CDRs into the single-chain antibody and/or CDR randomization prior to or concomitant with shuffling. The sequences of the synthetic CDR cassettes are selected by referring to known sequence data of human CDR and are selected in the discretion of the practitioner according to the following guidelines: synthetic CDRs will have at least 40 percent positional sequence identity to known CDR sequences, and preferably will have at least 50 to 70 percent positional sequence identity to known CDR sequences. For example, a collection of synthetic CDR sequences can be generated by synthesizing a collection of oligonucleotide sequences on the basis of naturally-occurring human CDR sequences listed in Kabat (Kabat et al, 1991); the pool (s) of synthetic CDR sequences are calculated to encode CDR peptide sequences having at least 40 percent sequence identity to at least one known naturally-occurring human CDR sequence. Alternatively, a collection of naturally-occurring CDR sequences may be compared to generate consensus sequences so that amino acids used at a residue position frequently (i.e., in at least 5 percent of known CDR sequences) are incorporated into the synthetic CDRs at the corresponding position(s). Typically, several (e.g., 3 to about 50) known CDR sequences are compared and observed natural sequence variations between the known CDRs are tabulated, and a collection of oligonucleotides encoding CDR peptide sequences

15

20

25

- encompassing all or most permutations of the observed natural sequence variations is synthesized. For example but not for limitation, if a collection of human VH CDR sequences have carboxy-terminal amino acids which are either Tyr, Val, Phe, or Asp, then the pool(s) of synthetic CDR oligonucleotide sequences are designed to allow the
- 5 carboxy-terminal CDR residue to be any of these amino acids. In some embodiments, residues other than those which naturally-occur at a residue position in the collection of CDR sequences are incorporated: conservative amino acid substitutions are frequently incorporated and up to 5 residue positions may be varied to incorporate non-conservative amino acid substitutions as compared to known naturally-occurring CDR sequences.
- 10 Such CDR sequences can be used in primary library members (prior to first round screening) and/or can be used to spike *in vitro* shuffling reactions of selected library member sequences. Construction of such pools of defined and/or degenerate sequences will be readily accomplished by those of ordinary skill in the art.
- 15 The collection of synthetic CDR sequences comprises at least one member that is not known to be a naturally-occurring CDR sequence. It is within the discretion of the practitioner to include or not include a portion of random or pseudorandom sequence corresponding to N region addition in the heavy chain CDR; the N region sequence ranges from 1 nucleotide to about 4 nucleotides occurring at V-D and D-J junctions. A
- 20 collection of synthetic heavy chain CDR sequences comprises at least about 100 unique CDR sequences, typically at least about 1,000 unique CDR sequences, preferably at least about 10,000 unique CDR sequences, frequently more than 50,000 unique CDR sequences; however, usually not more than about 1×10^6 unique CDR sequences are included in the collection, although occasionally 1×10^7 to 1×10^8 unique CDR
- 25 sequences are present, especially if conservative amino acid substitutions are permitted at positions where the conservative amino acid substituent is not present or is rare (i.e., less than 0.1 percent) in that position in naturally-occurring human CDRS. In general, the number of unique CDR sequences included in a library should not exceed the expected number of primary transformants in the library by more than a factor of 10. Such

single-chain antibodies generally bind of about at least 1×10^{-6} M, preferably with an affinity of about at least 5×10^7 M⁻¹, more preferably with an affinity of at least 1×10^8 M⁻¹ to 1×10^9 M⁻¹ or more, sometimes up to 1×10^{10} M⁻¹ or more. Frequently, the predetermined antigen is a human protein, such as for example a human cell surface antigen (e. g., CD4, CD8, IL-2 receptor, EGF receptor, PDGF receptor), other human biological macromolecule (e.g., thrombomodulin, protein C, carbohydrate antigen, sialyl Lewis antigen, Lselectin), or nonhuman disease associated macromolecule (e.g., bacterial LPS, virion capsid protein or envelope glycoprotein) and the like.

High affinity single-chain antibodies of the desired specificity can be engineered and expressed in a variety of systems. For example, scfv have been produced in plants (Firek et al, 1993) and can be readily made in prokaryotic systems (Owens and Young, 1994; Johnson and Bird, 1991). Furthermore, the single-chain antibodies can be used as a basis for constructing whole antibodies or various fragments thereof (Kettleborough et al, 1994). The variable region encoding sequence may be isolated (e.g., by PCR amplification or subcloning) and spliced to a sequence encoding a desired human constant region to encode a human sequence antibody more suitable for human therapeutic uses where immunogenicity is preferably minimized. The polynucleotide(s) having the resultant fully human encoding sequence(s) can be expressed in a host cell (e.g., from an expression vector in a mammalian cell) and purified for pharmaceutical formulation.

The DNA expression constructs will typically include an expression control DNA sequence operably linked to the coding sequences, including naturally-associated or heterologous promoter regions. Preferably, the expression control sequences will be eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells. Once the vector has been incorporated into the appropriate host, the host is maintained under conditions suitable for high level expression of the

nucleotide sequences, and the collection and purification of the mutant "engineered" antibodies.

As stated previously, the DNA sequences will be expressed in hosts after the
5 sequences have been operably linked to an expression control sequence (i.e., positioned
to ensure the transcription and translation of the structural gene). These expression
vectors are typically replicable in the host organisms either as episomes or as an integral
part of the host chromosomal DNA. Commonly, expression vectors will contain
selection markers, e.g., tetracycline or neomycin, to permit detection of those cells
10 transformed with the desired DNA sequences (see, e.g., USPN 4,704,362, which is
incorporated herein by reference).

In addition to eukaryotic microorganisms such as yeast, mammalian tissue cell
culture may also be used to produce the polypeptides of the present invention (see
15 Winnacker, 1987), which is incorporated herein by reference). Eukaryotic cells are
actually preferred, because a number of suitable host cell lines capable of secreting intact
immunoglobulins have been developed in the art, and include the CHO cell lines, various
COS cell lines, HeLa cells, and myeloma cell lines, but preferably transformed Bcells or
hybridomas. Expression vectors for these cells can include expression control sequences,
20 such as an origin of replication, a promoter, an enhancer (Queen et al, 1986), and
necessary processing information sites, such as ribosome binding sites, RNA splice sites,
polyadenylation sites, and transcriptional terminator sequences. Preferred expression
control sequences are promoters derived from immunoglobulin genes, cytomegalovirus,
SV40, Adenovirus, Bovine Papilloma Virus, and the like.

25

Eukaryotic DNA transcription can be increased by inserting an enhancer
sequence into the vector. Enhancers are cis-acting sequences of between 10 to 300 bp
that increase transcription by a promoter. Enhancers can effectively increase
transcription when either 5' or 3' to the transcription unit. They are also effective if

located within an intron or within the coding sequence itself. Typically, viral enhancers are used, including SV40 enhancers, cytomegalovirus enhancers, polyoma enhancers, and adenovirus enhancers. Enhancer sequences from mammalian systems are also commonly used, such as the mouse immunoglobulin heavy chain enhancer.

5

Mammalian expression vector systems will also typically include a selectable marker gene. Examples of suitable markers include, the dihydrofolate reductase gene (DHFR), the thymidine kinase gene (TK), or prokaryotic genes conferring drug resistance. The first two marker genes prefer the use of mutant cell lines that lack the
10 ability to grow without the addition of thymidine to the growth medium. Transformed cells can then be identified by their ability to grow on non-supplemented media. Examples of prokaryotic drug resistance genes useful as markers include genes conferring resistance to G418, mycophenolic acid and hygromycin.

15

The vectors containing the DNA segments of interest can be transferred into the host cell by well-known methods, depending on the type of cellular host. For example, calcium chloride transfection is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment, lipofection, or electroporation may be used for other cellular hosts. Other methods used to transform mammalian cells include the use of Polybrene,
20 protoplast fusion, liposomes, electroporation, and micro-injection (see, generally, Sambrook et al, 1982 and 1989).

Once expressed, the antibodies, individual mutated immunoglobulin chains, mutated antibody fragments, and other immunoglobulin polypeptides of the invention can
25 be purified according to standard procedures of the art, including ammonium sulfate precipitation, fraction column chromatography, gel electrophoresis and the like (see, generally, Scopes, 1982). Once purified, partially or to homogeneity as desired, the polypeptides may then be used therapeutically or in developing and performing assay

procedures, immunofluorescent stainings, and the like (see, generally, **Lefkovits** and **Pernis**, 1979 and 1981; **Lefkovits**, 1997).

5 The antibodies generated by the method of the present invention can be used for diagnosis and therapy. By way of illustration and not limitation, they can be used to treat cancer, autoimmune diseases, or viral infections. For treatment of cancer, the antibodies will typically bind to an antigen expressed preferentially on cancer cells, such as erbB-2, CEA, CD33, and many other antigens and binding members well known to those skilled in the art.

10

Two-Hybrid Based Screening Assays

Shuffling can also be used to recombinatorially diversify a pool of selected library
15 members obtained by screening a two-hybrid screening system to identify library members which bind a predetermined polypeptide sequence. The selected library members are pooled and shuffled by *in vitro* and/or *in vivo* recombination. The shuffled pool can then be screened in a yeast two hybrid system to select library members which bind said predetermined polypeptide sequence (e. g., and SH2 domain) or which bind an
20 alternate predetermined polypeptide sequence (e.g., an SH2 domain from another protein species).

An approach to identifying polypeptide sequences which bind to a predetermined polypeptide sequence has been to use a so-called "two-hybrid" system wherein the
25 predetermined polypeptide sequence is present in a fusion protein (**Chien** et al, 1991). This approach identifies protein-protein interactions *in vivo* through reconstitution of a transcriptional activator (**Fields** and **Song**, 1989), the yeast Gal4 transcription protein. Typically, the method is based on the properties of the yeast Gal4 protein, which consists of separable domains responsible for DNA-binding and transcriptional activation.

Polynucleotides encoding two hybrid proteins, one consisting of the yeast Gal4 DNA-binding domain fused to a polypeptide sequence of a known protein and the other consisting of the Gal4 activation domain fused to a polypeptide sequence of a second protein, are constructed and introduced into a yeast host cell. Intermolecular binding
5 between the two fusion proteins reconstitutes the Gal4 DNA-binding domain with the Gal4 activation domain, which leads to the transcriptional activation of a reporter gene (e.g., *lacZ*, *HIS3*) which is operably linked to a Gal4 binding site. Typically, the two-hybrid method is used to identify novel polypeptide sequences which interact with a known protein (Silver and Hunt, 1993; Durfee et al, 1993; Yang et al, 1992; Luban et
10 al, 1993; Hardy et al, 1992; Bartel et al, 1993; and Vojtek et al, 1993). However, variations of the two-hybrid method have been used to identify mutations of a known protein that affect its binding to a second known protein (Li and Fields, 1993; Lalo et al, 1993; Jackson et al, 1993; and Madura et al, 1993). Two-hybrid systems have also been used to identify interacting structural domains of two known proteins (Bardwell et al,
15 1993; Chakrabarty et al, 1992; Staudinger et al, 1993; and Milne and Weaver 1993) or domains responsible for oligomerization of a single protein (Iwabuchi et al, 1993; Bogerd et al, 1993). Variations of two-hybrid systems have been used to study the *in vivo* activity of a proteolytic enzyme (Dasmahapatra et al, 1992). Alternatively, an *E. coli*/BCCP interactive screening system (Germino et al, 1993; Guarente, 1993) can be
20 used to identify interacting protein sequences (i.e., protein sequences which heterodimerize or form higher order heteromultimers). Sequences selected by a two-hybrid system can be pooled and shuffled and introduced into a two-hybrid system for one or more subsequent rounds of screening to identify polypeptide sequences which bind to the hybrid containing the predetermined binding sequence. The sequences thus
25 identified can be compared to identify consensus sequence(s) and consensus sequence kernels.

In general, standard techniques of recombination DNA technology are described in various publications (e.g. Sambrook et al, 1989; Ausubel et al, 1987; and Berger and

Kimmel, 1987; each of which is incorporated herein in its entirety by reference. Polynucleotide modifying enzymes were used according to the manufacturer's recommendations. Oligonucleotides were synthesized on an Applied Biosystems Inc. Model 394 DNA synthesizer using ABI chemicals. If desired, PCR amplimers for
5 amplifying a predetermined DNA sequence may be selected at the discretion of the practitioner.

One microgram samples of template DNA are obtained and treated with U.V. light to cause the formation of dimers, including TT dimers, particularly purine dimers. U.V.
10 exposure is limited so that only a few photoproducts are generated per gene on the template DNA sample. Multiple samples are treated with U.V. light for varying periods of time to obtain template DNA samples with varying numbers of dimers from U.V. exposure.

15 A random priming kit which utilizes a non-proofreading polymease (for example, Prime-It II Random Primer Labeling kit by Stratagene Cloning Systems) is utilized to generate different size polynucleotides by priming at random sites on templates which are prepared by U.V. light (as described above) and extending along the templates. The priming protocols such as described in the Prime-It II Random Primer Labeling kit may
20 be utilized to extend the primers. The dimers formed by U.V. exposure serve as a roadblock for the extension by the non-proofreading polymerase. Thus, a pool of random size polynucleotides is present after extension with the random primers is finished.

The present invention is further directed to a method for generating a selected
25 mutant polynucleotide sequence (or a population of selected polynucleotide sequences) typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequences(s) possess at least one desired phenotypic characteristic (e.g., encodes a polypeptide, promotes transcription of linked polynucleotides, binds a protein, and the like) which can be selected for. One method for identifying hybrid polypeptides

that possess a desired structure or functional property, such as binding to a predetermined biological macromolecule (e.g., a receptor), involves the screening of a large library of polypeptides for individual library members which possess the desired structure or functional property conferred by the amino acid sequence of the polypeptide.

5

In one embodiment, the present invention provides a method for generating libraries of displayed polypeptides or displayed antibodies suitable for affinity interaction screening or phenotypic screening. The method comprises (1) obtaining a first plurality of selected library members comprising a displayed polypeptide or displayed antibody
10 and an associated polynucleotide encoding said displayed polypeptide or displayed antibody, and obtaining said associated polynucleotides or copies thereof wherein said associated polynucleotides comprise a region of substantially identical sequences, optimally introducing mutations into said polynucleotides or copies, (2) pooling the polynucleotides or copies, (3) producing smaller or shorter polynucleotides by
15 interrupting a random or particularized priming and synthesis process or an amplification process, and (4) performing amplification, preferably PCR amplification, and optionally mutagenesis to homologously recombine the newly synthesized polynucleotides.

It is a particularly preferred object of the invention to provide a process for
20 producing hybrid polynucleotides which express a useful hybrid polypeptide by a series of steps comprising:

(a) producing polynucleotides by interrupting a polynucleotide amplification or synthesis process with a means for blocking or interrupting the amplification or synthesis process and thus providing a plurality of smaller or shorter polynucleotides due
25 to the replication of the polynucleotide being in various stages of completion;

(b) adding to the resultant population of single- or double-stranded polynucleotides one or more single- or double-stranded oligonucleotides, wherein said added oligonucleotides comprise an area of identity in an area of heterology to one or more of the single- or double-stranded polynucleotides of the population;

(c) denaturing the resulting single- or double-stranded oligonucleotides to produce a mixture of single-stranded polynucleotides, optionally separating the shorter or smaller polynucleotides into pools of polynucleotides having various lengths and further optionally subjecting said polynucleotides to a PCR procedure to amplify one or more
5 oligonucleotides comprised by at least one of said polynucleotide pools;

(d) incubating a plurality of said polynucleotides or at least one pool of said polynucleotides with a polymerase under conditions which result in annealing of said single-stranded polynucleotides at regions of identity between the single-stranded polynucleotides and thus forming of a mutagenized double-stranded polynucleotide
10 chain;

(e) optionally repeating steps (c) and (d);

(f) expressing at least one hybrid polypeptide from said polynucleotide chain, or chains; and

(g) screening said at least one hybrid polypeptide for a useful activity.

15 In a preferred aspect of the invention, the means for blocking or interrupting the amplification or synthesis process is by utilization of uv light, DNA adducts, DNA binding proteins.

In one embodiment of the invention, the DNA adducts, or polynucleotides
20 comprising the DNA adducts, are removed from the polynucleotides or polynucleotide pool, such as by a process including heating the solution comprising the DNA fragments prior to further processing.

Having thus disclosed exemplary embodiments of the present invention, it should
25 be noted by those skilled in the art that the disclosures are exemplary only and that various other alternatives, adaptations and modifications may be made within the scope of the present invention. Accordingly, the present invention is not limited to the specific embodiments as illustrated herein.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following examples are to be considered illustrative and thus are not limiting of the remainder of the disclosure in any way whatsoever.

5

Example 1

Generation of Random Size Polynucleotides Using U.V. Induced Photoproducts

One microgram samples of template DNA are obtained and treated with U.V. light
10 to cause the formation of dimers, including TT dimers, particularly purine dimers. U.V. exposure is limited so that only a few photoproducts are generated per gene on the template DNA sample. Multiple samples are treated with U.V. light for varying periods of time to obtain template DNA samples with varying numbers of dimers from U.V. exposure.

15 A random priming kit which utilizes a non-proofreading polymerase (for example, Prime-It II Random Primer Labeling kit by Stratagene Cloning Systems) is utilized to generate different size polynucleotides by priming at random sites on templates which are prepared by U.V. light (as described above) and extending along the templates. The priming protocols such as described in the Prime-It II Random Primer
20 Labeling kit may be utilized to extend the primers. The dimers formed by U.V. exposure serve as a roadblock for the extension by the non-proofreading polymerase. Thus, a pool of random size polynucleotides is present after extension with the random primers is finished.

Example 2

Isolation of Random Size Polynucleotides

Polynucleotides of interest which are generated according to Example 1 are gel isolated on a 1.5% agarose gel. Polynucleotides in the 100-300 bp range are cut out of the gel and 3 volumes of 6 M NaI is added to the gel slice. The mixture is incubated at 50 °C for 10 minutes and 10 µl of glass milk (Bio 101) is added. The mixture is spun for 1 minute and the supernatant is decanted. The pellet is washed with 500 µl of Column Wash (Column Wash is 50% ethanol, 10mM Tris-HCl pH 7.5, 100 mM NaCl and 2.5 mM EDTA) and spin for 1 minute, after which the supernatant is decanted. The washing, spinning and decanting steps are then repeated. The glass milk pellet is resuspended in 20µl of H₂O and spun for 1 minute. DNA remains in the aqueous phase.

Example 3

Shuffling of Isolated Random Size 100-300bp Polynucleotides

The 100-300 bp polynucleotides obtained in Example 2 are recombined in an annealing mixture (0.2 mM each dNTP, 2.2 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl
5 ph 8.8, 0.1% Triton X-100, 0.3 μ; Taq DNA polymerase, 50 μl total volume) without adding primers. A Robocycler by Stratagene was used for the annealing step with the following program: 95 °C for 30 seconds, 25-50 cycles of [95 °C for 30 seconds, 50 - 60 °C (preferably 58 °C) for 30 seconds, and 72 °C for 30 seconds] and 5 minutes at 72 °C. Thus, the 100-300 bp polynucleotides combine to yield double-stranded polynucleotides
10 having a longer sequence. After separating out the reassembled double-stranded polynucleotides and denaturing them to form single stranded polynucleotides, the cycling is optionally again repeated with some samples utilizing the single strands as template and primer DNA and other samples utilizing random primers in addition to the single strands.

15

Example 4

Screening of Polypeptides from Shuffled Polynucleotides

The polynucleotides of Example 3 are separated and polypeptides are expressed therefrom. The original template DNA is utilized as a comparative control by obtaining
5 comparative polypeptides therefrom. The polypeptides obtained from the shuffled polynucleotides of Example 3 are screened for the activity of the polypeptides obtained from the original template and compared with the activity levels of the control. The shuffled polynucleotides coding for interesting polypeptides discovered during screening are compared further for secondary desirable traits. Some shuffled polynucleotides
10 corresponding to less interesting screened polypeptides are subjected to reshuffling.

Example 5

Directed Evolution an Enzyme by Saturation Mutagenesis

Site-Saturation Mutagenesis: To accomplish site-saturation mutagenesis every residue
5 (316) of a dehalogenase enzyme was converted into all 20 amino acids by site directed
mutagenesis using 32-fold degenerate oligonucleotide primers, as follows:

1. A culture of the dehalogenase expression construct was grown and a preparation of
the plasmid was made
- 10 2. Primers were made to randomize each codon - they have the common structure
 $X_{20}NN(G/T)X_{20}$
3. A reaction mix of 25 ul was prepared containing ~50 ng of plasmid template, 125 ng
of each primer, 1X native Pfu buffer, 200 uM each dNTP and 2.5 U native Pfu DNA
polymerase
- 15 4. The reaction was cycled in a Robo96 Gradient Cycler as follows:
Initial denaturation at 95°C for 1 min
20 cycles of 95°C for 45 sec, 53°C for 1 min and 72°C for 11 min
Final elongation step of 72°C for 10 min
5. The reaction mix was digested with 10 U of DpnI at 37°C for 1 hour to digest the
20 methylated template DNA
6. Two ul of the reaction mix were used to transform 50 ul of XL1-Blue MRF' cells and
the entire transformation mix was plated on a large LB-Amp-Met plate yielding 200-
1000 colonies
7. Individual colonies were toothpicked into the wells of 96-well microtiter plates
25 containing LB-Amp-IPTG and grown overnight
8. The clones on these plates were assayed the following day

Screening: Approximately 200 clones of mutants for each position were grown in liquid media (384 well microtiter plates) and screened as follows:

- 5 1. Overnight cultures in 384-well plates were centrifuged and the media removed. To each well was added 0.06 mL 1 mM Tris/SO₄²⁻ pH 7.8.
2. Made 2 assay plates from each parent growth plate consisting of 0.02 mL cell suspension.
- 10 3. One assay plate was placed at room temperature and the other at elevated temperature (initial screen used 55°C) for a period of time (initially 30 minutes).
4. After the prescribed time 0.08 mL room temperature substrate (TCP saturated 1 mM Tris/SO₄²⁻ pH 7.8 with 1.5 mM NaN₃ and 0.1 mM bromothymol blue) was added to each well.
- 15 5. Measurements at 620 nm were taken at various time points to generate a progress curve for each well.
6. Data were analyzed and the kinetics of the cells heated to those not heated were compared. Each plate contained 1-2 columns (24 wells) of unmutated 20F12 controls.
- 20 7. Wells that appeared to have improved stability were re-grown and tested under the same conditions.

Following this procedure nine single site mutations appeared to confer increased thermal stability on the enzyme. Sequence analysis was performed to determine of the
25 exact amino acid changes at each position that were specifically responsible for the improvement. In sum, the improvement was conferred at 7 sites by one amino acid change alone, at an eighth site by each of two amino acid changes, and at a ninth site by each of three amino acid changes. Several mutants were then made each having a

plurality of these nine beneficial site mutations in combination; of these two mutants proved superior to all the other mutants, including those with single point mutations.

Example 6

Direct expression cloning using end-selection

An esterase gene was amplified using 5'phosphorylated primers in a standard
5 PCR reaction (10 ng template; PCR conditions: 3' 94 C; [1' 94 C; 1' 50 C; 1'30" 68 C] x
30; 10' 68 C.

Forward Primer = 9511TopF

(CTAGAAGGGAGGAGAATTACATGAAGCGGCTTTTAGCCC)

10 Reverse Primer = 9511TopR (AGCTAAGGGTCAAGGCCGCACCCGAGG)

The resulting PCR product (ca.1000 bp) was gel purified and quantified.

A vector for expression cloning, pASK3 (Institut fuer Bioanalytik, Goettingen,
Germany), was cut with *Xba* I and *Bgl* II and dephosphorylated with CIP.

15

0.5 pmoles Vaccina Topoisomerase I (Invitrogen, Carlsbad, CA) was added to 60
ng (ca. 0.1 pmole) purified PCR product for 5' 37 C in buffer NEB I (New England
Biolabs, Beverly, MA) in 5 µl total volume.

The topogated PCR product was cloned into the vector pASK3 (5 µl, ca. 200 ng in NEB
20 I) for 5' at room temperature.

This mixture was dialyzed against H₂O for 30'.

2 µl were used for electroporation of DH10B cells (Gibco BRL, Gaithersburg, MD).

Efficiency: Based on the actual clone numbers this method can produce 2×10^6
25 clones per µg vector. All tested recombinants showed esterase activity after induction
with anhydrotetracycline.

Example 7

Dehalogenase Thermal Stability

5 This invention provides that a desirable property to be generated by directed evolution is exemplified in a limiting fashion by an improved residual activity (e.g. an enzymatic activity, an immunoreactivity, an antibiotic activity, etc.) of a molecule upon
10 subjection to altered environment, including what may be considered a harsh environment, for a specified time. Such a harsh environment may comprise any combination of the following (iteratively or not, and in any order or permutation): an elevated temperature
(including a temperature that may cause denaturation of a working enzyme), a decreased temperature, an elevated salinity, a decreased salinity, an elevated pH, a decreased pH, an elevated pressure, a decreased pressure, and an change in exposure to a radiation source
(including uv radiation, visible light, as well as the entire electromagnetic spectrum).

15 The following example shows an application of directed evolution to evolve the ability of an enzyme to regain &/or retain activity upon exposure to an elevated temperature.

Every residue (316) of a dehalogenase enzyme was converted into all 20 amino acids by site directed mutagenesis using 32-fold degenerate oligonucleotide primers. These
20 mutations were introduced into the already rate-improved variant Dhla 20F12. Approximately 200 clones of each position were grown in liquid media (384 well microtiter plates) to be screened. The screening procedure was as follows:

1. Overnight cultures in 384-well plates were centrifuged and the media removed.
25 To each well was added 0.06 mL 1 mM Tris/SO₄²⁻ pH 7.8.
2. The robot made 2 assay plates from each parent growth plate consisting of 0.02 mL cell suspension.
3. One assay plate was placed at room temperature and the other at elevated temperature (initial screen used 55°C) for a period of time (initially 30 minutes).

4. After the prescribed time 0.08 mL room temperature substrate (TCP saturated 1 mM Tris/SO₄²⁻ pH 7.8 with 1.5 mM NaN₃ and 0.1 mM bromothymol blue) was added to each well. TCP = trichloropropane.
5. Measurements at 620 nm were taken at various time points to generate a progress curve for each well.
6. Data were analyzed and the kinetics of the cells heated to those not heated were compared. Each plate contained 1-2 columns (24 wells) of un-mutated 20F12 controls.
7. Wells that appeared to have improved stability were regrown and tested under the same conditions.

Following this procedure nine single site mutations appeared to confer increased thermal stability on Dhla-20F12. Sequence analysis showed that the following changes were beneficial:

15

D89G

F91S

T159L

G189Q, G189V

20

I220L

N238T

W251Y

P302A, P302L, P302S, P302K

P302R/S306R

25

Only two sites (189 and 302) had more than one substitution. The first 5 on the list were combined (using G189Q) into a single gene (this mutant is referred to as "Dhla5"). All changes but S306R were incorporated into another variant referred to as Dhla8.

Thermal stability was assessed by incubating the enzyme at the elevated temperature (55°C and 80°C) for some period of time and activity assay at 30°C. Initial rates were plotted vs. time at the higher temperature. The enzyme was in 50 mM Tris/SO₄ pH 7.8 for both the incubation and the assay. Product (Cl⁻) was detected by a standard method
5 using Fe(NO₃)₃ and HgSCN. Dhla 20F12 was used as the *de facto* wild type. The apparent half-life (T_{1/2}) was calculated by fitting the data to an exponential decay function.

3. LITERATURE CITED

Unless otherwise indicated, all references cited herein (supra and infra) are incorporated by reference in their entirety.

5

Barret AJ, et al., eds.: Enzyme Nomenclature: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology. San Diego: Academic Press, Inc., 1992.

- 10 Boyce COL, ed.: Novo's Handbook of Practical Biotechnology. 2nd ed. Bagsvaerd, Denmark, 1986.

Drauz K, Waldman H, eds.: Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook. Vol. 1. New York: VCH Publishers, 1995.

15

Drauz K, Waldman H, eds.: Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook. Vol. 2. New York: VCH Publishers, 1995.

- 20 Foster GD, Taylor SC, eds.: Plant Virology Protocols: From Virus Isolation to Transgenic Resistance. Methods in Molecular Biology, Vol. 81. New Jersey: Humana Press Inc., 1998.

Franks F, ed.: Protein Biotechnology: Isolation, Characterization, and Stabilization. New Jersey: Humana Press Inc., 1993.

25

Godfrey T, West S, eds.: Industrial Enzymology. 2nd ed. London: Macmillan Press Ltd, 1996.

Gottschalk G: Bacterial Metabolism. 2nd ed. New York: Springer-Verlag Inc., 1986.

Gresshoff PM, ed.: Technology Transfer of Plant Biotechnology. Current Topics in Plant Molecular Biology. Boca Raton: CRC Press, 1997.

- 5 Griffin HG, Griffin AM, eds.: PCR Technology: Current Innovations. Boca Raton: CRC Press, Inc., 1994.

Hansen G, Chilton MD: Lessons in gene transfer to plants by a gifted microbe. *Curr Top Microbiol Immunol* 240:21-57, 1999.

10

Hartmann HT, et al.: Plant Propagation: Principles and Practices. 6th ed. New Jersey: Prentice Hall, Inc., 1997.

Perun TJ, Propst CL, eds.: Computer-Aided Drug Design: Methods and Applications.

- 15 New York: Marcel Dekker, Inc., 1989.

Owen MRL, Pen J: Transgenic Plants: A Production System for Industrial and Pharmaceutical Proteins. Chichester: John Wiley & Sons, 1996.

- 20 Segel IH: Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems. New York: John Wiley & Sons, Inc., 1993.

White JS, White DC: Source Book of Enzymes. Boca Raton: CRC Press, 1997.

- 25 Wong CH, Whitesides GM: Enzymes in Synthetic Organic Chemistry. Vol. 12. New York: Elsevier Science Publications, 1995.

WO 97/35966; Filed Mar. 20, 1997, Published Oct. 2, 1997. Minshull J, Stemmer WP: Methods and Compositions for Cellular and Metabolic Engineering.

WO 98/31837; Filed Jan. 16, 1998, Published July 23, 1998. Delcardayre SB, Tobin MB, Stemmer WP, Minshull, J: Evolution of Whole Cells and Organisms by Recursive Sequence Recombination.

5

WO 98/37223; Filed Feb. 18 1998, Published Aug. 27, 1998. Pang SZ, Gonsalves D, Jan FJ: DNA Construct to Confer Multiple Traits on Plants.

Alting-Mecs MA and Short JM: Polycos vectors: a system for packaging filamentous
10 phage and phagemid vectors using lambda phage packaging extracts. *Gene* 137:1, 93-100, 1993.

Arkin AP and Youvan DC: An algorithm for protein engineering: simulations of recursive ensemble mutagenesis. *Proc Natl Acad Sci USA* 89(16):7811-7815, (Aug 15)
15 1992.

Arnold FH: Protein engineering for unusual environments. *Current Opinion in Biotechnology* 4(4):450-455, 1993.

20 Ausubel FM, et al Editors. Current Protocols in Molecular Biology, Vols. 1 and 2 and supplements. (a.k.a. "The Red Book") Greene Publishing Assoc., Brooklyn, NY, ©1987.

Ausubel FM, et al Editors. Current Protocols in Molecular Biology, Vols. 1 and 2 and supplements. (a.k.a. "The Red Book") Greene Publishing Assoc., Brooklyn, NY, ©1989.

25

Ausubel FM, et al Editors. Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology. Greene Publishing Assoc., Brooklyn, NY, ©1989.

Ausubel FM, et al Editors. Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology, 2nd Edition. Greene Publishing Assoc., Brooklyn, NY, ©1992.

- 5 Barbas CF 3d, Bain JD, Hoekstra DM, Lerner RA: Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc Natl Acad Sci USA* 89(10):4457-4461, 1992.
- 10 Bardwell AJ, Bardwell L, Johnson DK, Friedberg EC: Yeast DNA recombination and repair proteins Rad1 and Rad10 constitute a complex in vivo mediated by localized hydrophobic domains. *Mol Microbiol* 8(6):1177-1188, 1993.
- 15 Bartel P, Chien CT, Sternglanz R, Fields S: Elimination of false positives that arise in using the two-hybrid system. *Biotechniques* 14(6):920-924, 1993.
- Beaudry AA and Joyce GF: Directed evolution of an RNA enzyme. *Science* 257(5070):635-641, 1992.
- 20 Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques. Academic Press, Inc., San Diego, CA, ©1987. (Cumulative Subject Index: Volumes 135-139, 141-167, 1990, 272 pp.)
- 25 Bevan M: Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Research* 12(22):8711-21, 1984.
- Biocca S, Pierandrei-Amaldi P, Cattaneo A: Intracellular expression of anti-p21ras single chain Fv fragments inhibits meiotic maturation of xenopus oocytes. *Biochem Biophys Res Commun* 197(2):422-427, 1993.

- Bird et al. *Plant Mol Biol* 11:651, 1988..
- 5 Bogerd HP, Fridell RA, Blair WS, Cullen BR: Genetic evidence that the Tat proteins of human immunodeficiency virus types 1 and 2 can multimerize in the eukaryotic cell nucleus. *J Virol* 67(8):5030-5034, 1993.
- Brederode FT, Koper-Zawrthoff EC, Bol JF: Complete nucleotide sequence of alfalfa mosaic virus RNA 4. *Nucleic Acids Research* 8(10):2213-23, 1980.
- 10 Breitling F, Dubel S, Seehaus T, Klewinghaus I, Little M: A surface expression vector for antibody screening. *Gene* 104(2):147-153, 1991.
- Brown NL, Smith M: Cleavage specificity of the restriction endonuclease isolated from *Haemophilus gallinarum* (Hga I). *Proc Natl Acad Sci U S A* 74(8):3213-6, (Aug) 1977.
- 15 Burton DR, Barbas CF 3d, Persson MA, Koenig S, Chanock RM, Lerner RA: A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. *Proc Natl Acad Sci U S A* 88(22):10134-7, (Nov 15) 1991.
- 20 Caldwell RC and Joyce GF: Randomization of genes by PCR mutagenesis. *PCR Methods Appl* 2(10):28-33, 1992.
- 25 Caton AJ and Koprowski H: Influenza virus hemagglutinin-specific antibodies isolated from a combinatorial expression library are closely related to the immune response of the donor. *Proc Natl Acad Sci USA* 87(16):6450-6454, 1990.

- Chakraborty T, Martin JF, Olson EN: Analysis of the oligomerization of myogenin and E2A products in vivo using a two-hybrid assay system. *J Biol Chem* 267(25):17498-501, 1992.
- 5 Chang CN, Landolfi NF, Queen C: Expression of antibody Fab domains on bacteriophage surfaces. Potential use for antibody selection. *J Immunol* 147(10):3610-4, (Nov 15) 1991.
- Chaudhary VK, Batra JK, Gallo MG, Willingham MC, FitzGerald DJ, Pastan I: A rapid
10 method of cloning functional variable-region antibody genes in *Escherichia coli* as single-chain immunotoxins. *Proc Natl Acad Sci USA* 87(3):1066-1070, 1990.
- Chien CT, Bartel PL, Sternglanz R, Fields S: The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl*
15 *Acad Sci USA* 88(21):9578-9582, 1991.
- Chiswell DJ, McCafferty J: Phage antibodies: will new 'coliclonal' antibodies replace monoclonal antibodies? *Trends Biotechnol* 10(3):80-84, 1992.
- 20 Chothia C and Lesk AM: Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 196(4):901-917, 1987.
- Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, et al: Conformations of immunoglobulin hypervariable
25 regions. *Nature* 342(6252):877-883, 1989.
- Clackson T, Hoogenboom HR, Griffiths AD, Winter G: Making antibody fragments using phage display libraries. *Nature* 352(6336):624-628, 1991.

Conrad M, Topal MD: DNA and spermidine provide a switch mechanism to regulate the activity of restriction enzyme *Nae I*. *Proc Natl Acad Sci USA* 86(24):9707-11, (Dec) 1989.

- 5 Coruzzi G, Broglie R, Edwards C, Chua NH: Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. *EMBO J* 3(8):1671-9, 1984.

- 10 Dasmahapatra B, DiDomenico B, Dwyer S, Ma J, Sadowski I, Schwartz J: A genetic system for studying the activity of a proteolytic enzyme. *Proc Natl Acad Sci USA* 89(9):4159-4162, 1992.

Davis LG, Dibner MD, Battey JF. Basic Methods in Molecular Biology. Elsevier, New York, NY, ©1986.

15

Delegrave S and Youvan DC. *Biotechnology Research* 11:1548-1552, 1993.

DeLong EF, Wu KY, Prezelin BB, Jovine RV: High abundance of Archaea in Antarctic marine picoplankton. *Nature* 371(6499):695-697, 1994.

20

Deng SJ, MacKenzie CR, Sadowska J, Michniewicz J, Young NM, Bundle Dr, Narang SA: Selection of antibody single-chain variable fragments with improved carbohydrate binding by phage display. *J Biol Chem* 269(13):9533-9538, 1994.

- 25 Duan L, Bagasra O, Laughlin MA, Oakes JW, Pomerantz RJ: Potent inhibition of human immunodeficiency virus type 1 replication by an intracellular anti-Rev single-chain antibody. *Proc Natl Acad Sci USA* 91(11):5075-5079, 1994.

Durfee T, Becherer K, Chen PL, Yeh SH, Yang Y, Kilburn AE, Lee WH, Elledge SJ: The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev* 7(4):555-569, 1993.

- 5 Ellington AD and Szostak JW: In vitro selection of RNA molecules that bind specific ligands. *Nature* 346(6287):818-822, 1990.

Fields S and Song O: A novel genetic system to detect protein-protein interactions. *Nature* 340(6230):245-246, 1989.

10

Firek S, Draper J, Owen MR, Gandeche A, Cockburn B, Whitlam GC: Secretion of a functional single-chain Fv protein in transgenic tobacco plants and cell suspension cultures. *Plant Mol Biol* 23(4):861-870, 1993.

- 15 Forsblom S, Rigler R, Ehrenberg M, Philipson L: Kinetic studies on the cleavage of adenovirus DNA by restriction endonuclease Eco RI. *Nucleic Acids Res* 3(12):3255-69, (Dec) 1976.

- Germino FJ, Wang ZX, Weissman SM: Screening for in vivo protein-protein
20 interactions. *Proc Natl Acad Sci USA* 90(3):933-937, 1993.

Gingeras TR, Brooks JE: Cloned restriction/modification system from *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 80(2):402-6, 1983 (Jan).

- 25 Gluzman Y: SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* 23(1):175-182, 1981.

Gruber M, Schodin BA, Wilson ER, Kranz DM: Efficient tumor cell lysis mediated by a bispecific single chain antibody expressed in *Escherichia coli*. *J Immunol* 152(11):5368-5374, 1994.

- 5 Guarente L: Strategies for the identification of interacting proteins. *Proc Natl Acad Sci USA* 90(5):1639-1641, 1993.

Guilley H, Dudley RK, Jonard G, Balazs E, Richards KE: Transcription of Cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of
10 transcripts. *Cell* 30(3):763-73, 1982.

Hardy CF, Sussel L, Shore D: A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes Dev* 6(5):801-814, 1992.

- 15 Hawkins RE and Winter G: Cell selection strategies for making antibodies from variable gene libraries: trapping the memory pool. *Eur J Immunol* 22(3):867-870, 1992.

Holvoet P, Laroche Y, Lijnen HR, Van Hoef B, Brouwers E, De Cock F, Lauwereys M, Gansemans Y, Collen D: Biochemical characterization of single-chain chimeric
20 plasminogen activators consisting of a single-chain Fv fragment of a fibrin-specific antibody and single-chain urokinase. *Eur J Biochem* 210(3):945-952, 1992.

Honjo T, Alt FW, Rabbitts TH (eds): *Immunoglobulin genes*. Academic Press: San Diego, CA, pp. 361-368, ©1989.

25

Hoogenboom HR, Griffiths AD, Johnson KS, Chiswell DJ, Judson P, Winter G: Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res* 19(15):4133-4137, 1991.

Huse WD, Sastry L, Iverson SA, Kang AS, Alting-Mees M, Burton DR, Benkovic SJ, Lerner RA: Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* 246(4935):1275-1281, 1989.

- 5 Huston JS, Levinson D, Mudgett-Hunter M, Tai MS, Novotney J, Margolies MN, Ridge RJ, Brucoleri RE, Haber E, Crea R, et al: Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc Natl Acad Sci USA* 85(16):5879-5883, 1988.

- 10 Iwabuchi K, Li B, Bartel P, Fields S: Use of the two-hybrid system to identify the domain of p53 involved in oligomerization. *Oncogene* 8(6):1693-1696, 1993.

Jackson AL, Pahl PM, Harrison K, Rosamond J, Sclafani RA: Cell cycle regulation of the yeast Cdc7 protein kinase by association with the Dbf4 protein. *Mol Cell Biol*

- 15 13(5):2899-2908, 1993.

Johnson S and Bird RE: *Methods Enzymol* 203:88, 1991.

- Kabat et al: Sequences of Proteins of Immunological Interest, 4th Ed. U.S. Department
20 of Health and Human Services, Bethesda, MD (1987)

Kang AS, Barbas CF, Janda KD, Benkovic SJ, Lerner RA: Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc Natl Acad Sci USA* 88(10):4363-4366, 1991.

25

Kettleborough CA, Ansell KH, Allen RW, Rosell-Vives E, Gussow DH, Bendig MM: Isolation of tumor cell-specific single-chain Fv from immunized mice using phage-antibody libraries and the re-construction of whole antibodies from these antibody fragments. *Eur J Immunol* 24(4):952-958, 1994.

Kruger DH, Barcak GJ, Reuter M, Smith HO: EcoRII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res* 16(9):3997-4008, (May 11) 1988.

- 5 Lalo D, Carles C, Sentenac A, Thuriaux P: Interactions between three common subunits of yeast RNA polymerases I and III. *Proc Natl Acad Sci USA* 90(12):5524-5528, 1993.

Laskowski M Sr: Purification and properties of venom phosphodiesterase. *Methods Enzymol* 65(1):276-84, 1980.

10

Lefkovits I and Pernis B, Editors. Immunological Methods, Vols. I and II. Academic Press, New York, NY. Also Vol. III published in Orlando and Vol. IV published in San Diego. ©1979-.

- 15 Ivan Lefkovits, Editor. Immunology methods manual : the comprehensive sourcebook of techniques. Academic Press, San Diego, ©1997.

Lerner RA, Kang AS, Bain JD, Burton DR, Barbas CF 3d: Antibodies without immunization. *Science* 258(5086):1313-1314, 1992.

20

Leung, D.W., et al, *Technique*, 1:11-15, 1989.

Li B and Fields S: Identification of mutations in p53 that affect its binding to SV40 large T antigen by using the yeast two-hybrid system. *FASEB J* 7(10):957-963, 1993.

25

Lilley GG, Doelzal O, Hillyard CJ, Bernard C, Hudson PJ: Recombinant single-chain antibody peptide conjugates expressed in Escherichia coli for the rapid diagnosis of HIV. *J Immunol Methods* 171(2):211-226, 1994.

- 30 Lowman HB, Bass SH, Simpson N, Wells JA: Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* 30(45):10832-10838, 1991.

Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP: Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* 73(6):1067-1078, 1993.

- 5 Madura K, Dohmen RJ, Varshavsky A: N-recognin/Ubc2 interactions in the N-end rule pathway. *J Biol Chem* 268(16):12046-54, (Jun 5) 1993.

Marks JD, Hoogenboom HR, Bonnert TP, McCafferty J, Griffiths AD, Winter G: Bypassing immunization. Human antibodies from V-gene libraries displayed on phage. *J Mol Biol* 222(3):581-597, 1991.

Marks JD, Griffiths AD, Malmqvist M, Clackson TP, Bye JM, Winter G: Bypassing immunization: building high affinity human antibodies by chain shuffling. *Biotechnology (N Y)* 10(7):779-783, 1992.

15 Marks JD, Hoogenboom HR, Griffiths AD, Winter G: Molecular evolution of proteins on filamentous phage. Mimicking the strategy of the immune system. *J Biol Chem* 267(23):16007-16010, 1992.

20 Maxam AM, Gilbert W: Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* 65(1):499-560, 1980.

McCafferty J, Griffiths AD, Winter G, Chiswell DJ: Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348(6301):552-554, 1990.

25 Miller JH. A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria (see inclusively p. 445). Cold Spring Harbor Laboratory Press, Plainview, NY, ©1992.

- Milne GT and Weaver DT: Dominant negative alleles of RAD52 reveal a DNA repair/recombination complex including Rad51 and Rad52. *Genes Dev* 7(9):1755-1765, 1993.
- Mullinax RL, Gross EA, Amberg JR, Hay BN, Hogrefe HH, Kubitiz MM, Greener A,
5 Alting-Mees M, Ardourel D, Short JM, et al: Identification of human antibody fragment clones specific for tetanus toxoid in a bacteriophage lambda immunoexpression library. *Proc natl Acad Sci USA* 87(20):8095-9099, 1990.
- Nath K, Azzolina BA: in *Gene Amplification and Analysis* (ed. Chirikjian JG), vol. 1, p.
10 113, Elsevier North Holland, Inc., New York, New York, ©1981.
- Needleman SB and Wunsch CD: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48(3):443-453, 1970.
- 15 Nelson M, Christ C, Schildkraut I: Alteration of apparent restriction endonuclease recognition specificities by DNA methylases. *Nucleic Acids Res* 12(13):5165-73, 1984 (Jul 11).
- Nicholls PJ, Johnson VG, Andrew SM, Hoogenboom HR, Raus JC, Youle RJ:
20 Characterization of single-chain antibody (sFv)-toxin fusion proteins produced in vitro in rabbit reticulocyte lysate. *J Biol Chem* 268(7):5302-5308, 1993.
- Oller AR, Vanden Broek W, Conrad M, Topal MD: Ability of DNA and spermidine to affect the activity of restriction endonucleases from several bacterial species.
25 *Biochemistry* 30(9):2543-9, (Mar 5) 1991.
- Owens RJ and Young RJ: The genetic engineering of monoclonal antibodies. *J Immunol Methods* 168(2):149-165, 1994.

Pearson WR and Lipman DJ: Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85(8):2444-2448, 1988.

Pein CD, Reuter M, Meisel A, Cech D, Kruger DH: Activation of restriction
5 endonuclease EcoRII does not depend on the cleavage of stimulator DNA. *Nucleic Acids Res* 19(19):5139-42, (Oct 11) 1991.

Persson MA, Caothien RH, Burton DR: Generation of diverse high-affinity human
monoclonal antibodies by repertoire cloning. *Proc Natl Acad Sci USA* 88(6):2432-2436,
10 1991.

Queen C, Foster J, Stauber C, Stafford J: Cell-type specific regulation of a kappa
immunoglobulin gene by promoter and enhance elements. *Immunol Rev* 89:49-68, 1986.

15 Qiang BQ, McClelland M, Poddar S, Spokauskas A, Nelson M: The apparent specificity
of NotI (5'-GCGGCCGC-3') is enhanced by M.FnuDII or M.BepI methyltransferases (5'-
mCGCG-3'): cutting bacterial chromosomes into a few large pieces. *Gene* 88(1):101-5,
(Mar 30) 1990.

20 Raleigh EA, Wilson G: Escherichia coli K-12 restricts DNA containing 5-
methylcytosine. *Proc Natl Acad Sci USA* 83(23):9070-4, (Dec) 1986.

Reidhaar-Olson JF and Sauer RT: Combinatorial cassette mutagenesis as a probe of the
informational content of protein sequences. *Science* 241(4861):53-57, 1988.

25 Riechmann L and Weill M: Phage display and selection of a site-directed randomized
single-chain antibody Fv fragment for its affinity improvement. *Biochemistry*
32(34):8848-8855, 1993.

Roberts RJ, Macelis D: REBASE--restriction enzymes and methylases. *Nucleic Acids Res* 24(1):223-35, (Jan 1) 1996.

5 Ryan AJ, Royal CL, Hutchinson J, Shaw CH: Genomic sequence of a 12S seed storage protein from oilseed rape (*Brassica napus* c.v. jet neuf). *Nucl Acids Res* 17(9):3584, 1989.

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ©1982.

10

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ©1989.

15 Scopes RK. Protein Purification: Principles and Practice. Springer-Verlag, New York, NY, © 1982.

Silver SC and Hunt SW 3d: Techniques for cloning cDNAs encoding interactive transcriptional regulatory proteins. *Mol Biol Rep* 17(3):155-165, 1993.

20 Smith TF, Waterman MS. *Adv Appl Math* 2: 482-end of article, 1981.

Smith TF, Waterman MS: Overlapping genes and information theory. *J Theor Biol* 91(2):379-80, (Jul 21) 1981.

25 Smith TF, Waterman MS: Identification of common molecular subsequences. *J Mol Biol* 147(1):195-7, (Mar 25) 1981.

Smith TF, Waterman MS, Fitch WM: Comparative biosequence metrics. *J Mol Evol* 18(1):38-46, 1981.

30

Staudinger J, Perry M, Elledge SJ, Olson EN: Interactions among vertebrate helix-loop-helix proteins in yeast using the two-hybrid system. *J Biol Chem* 268(7):4608-4611, 1993.

- 5 Stemmer WP, Morris SK, Wilson BS: Selection of an active single chain Fv antibody from a protein linker library prepared by enzymatic inverse PCR. *Biotechniques* 14(2):256-265, 1993.

- Stemmer WP: DNA shuffling by random fragmentation and reassembly: in vitro
10 recombination for molecular evolution. *Proc Natl Acad Sci USA* 91(22):10747-10751, 1994.

- Sun D, Hurley LH: Effect of the (+)-CC-1065-(N3-adenine)DNA adduct on in vitro DNA synthesis mediated by Escherichia coli DNA polymerase. *Biochemistry* 31:10,
15 2822-9, (Mar 17) 1992,

Tague BW, Dickinson CD, Chrispeels MJ: A short domain of the plant vacuolar protein phytohemagglutinin targets invertase to the yeast vacuole. *Plant Cell* 2(6):533-46, (June) 1990.

20

Takahashi N, Kobayashi I: Evidence for the double-strand break repair model of bacteriophage lambda recombination. *Proc Natl Acad Sci U S A* 87(7):2790-4, (Apr) 1990.

- 25 Thiesen HJ and Bach C: Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res* 18(11):3203-3209, 1990.

- Thomas M, Davis RW: Studies on the cleavage of bacteriophage lambda DNA with
30 EcoRI Restriction endonuclease. *J Mol Biol* 91(3):315-28, (Jan 25) 1975.

- Tingey SV, Walker EL, Corruzzi GM: Glutamine synthetase genes of pea encode distinct polypeptides which are differentially expressed in leaves, roots and nodules. *EMBO J* 6(1):1-9, 1987.
- 5 Topal MD, Thresher RJ, Conrad M, Griffith J: NaeI endonuclease binding to pBR322 DNA induces looping. *Biochemistry* 30(7):2006-10, (Feb. 19) 1991.
- 10 Tramontano A, Chothia C, Lesk AM: Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J Mol Biol* 215(1):175-182, 1990.
- Tuerk C and Gold L: Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249(4968):505-510, 1990.
- 15 van de Poll ML, Lafleur MV, van Gog F, Vrieling H, Meerman JH: N-acetylated and deacetylated 4'-fluoro-4-aminobiphenyl and 4-aminobiphenyl adducts differ in their ability to inhibit DNA replication of single-stranded M13 in vitro and of single-stranded phi X174 in Escherichia coli. *Carcinogenesis* 13(5):751-8, (May) 1992.
- 20 Vojtek AB, Hollenberg SM, Cooper JA: Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* 74(1):205-214, 1993.
- 25 Wenzler H, Mignery G, Fisher L, Park W: Sucrose-regulated expression of a chimeric potato tuber gene in leaves of transgenic tobacco plants. *Plant Mol Biol* 13(4):347-54, 1989.
- Williams and Barclay, in Immunoglobulin Genes, The Immunoglobulin Gene Superfamily

Winnacker EL. From Genes to Clones: Introduction to Gene Technology. VCH Publishers, New York, NY, ©1987.

- 5 Winter G and Milstein C: Man-made antibodies. *Nature* 349(6307):293-299, 1991.

Yang X, Hubbard EJ, Carlson M: A protein kinase substrate identified by the two-hybrid system. *Science* 257(5070):680-2, (Jul 31) 1992.

10

USPN 4,683,195; Filed Feb. 7, 1986, Issued Jul 28, 1987. Mullis KB, Erlich HA, Arnheim N, Horn GT, Saiki RK, Scharf SJ: Process for Amplifying, Detecting, and/or Cloning Nucleic Acid Sequences.

- 15 USPN 4,683,202; Filed Oct. 25, 1985, Issued Jul. 28, 1987. Mullis KB: Process for Amplifying Nucleic Acid Sequences.

USPN 4,704,362; Filed Nov. 5, 1979, Issued Nov. 3, 1987. Itakura K, Riggs AD: Recombinant Cloning Vehicle Microbial Polypeptide Expression.

20

WO 88/08453; Filed Apr. 14, 1988, Published Nov. 3, 1988. Alakhov JB, Baranov, VI, Ovodov SJ, Ryabova LA, Spirin AS: Method of Obtaining Polypeptides in Cell-Free Translation System.

- 25 WO 90/05785; Filed Nov. 15, 1989, Published May 31, 1990. Schultz P: Method for Site-Specifically Incorporating Unnatural Amino Acids into Proteins.

WO 90/07003; Filed Jan. 27, 1989, Published June 28, 1990. Baranov VI, Morozov IJ, Spirin AS: Method for Preparative Expression of Genes in a Cell-free System of

- 30 Conjugated Transcription/translation.

WO 91/02076; Filed June 14, 1990, Published Feb. 21, 1991. Baranov VI, Ryabova LA, Yarchuk OB, Spirin AS: Method for Obtaining Polypeptides in a Cell-free System.

- 5 WO 91/05058; Filed Oct. 5, 1989, Published Apr. 18, 1991. Kawasaki G: Cell-free Synthesis and Isolation of Novel Genes and Polypeptides.

WO 91/17271; Filed May 1, 1990, Published Nov. 14, 1991. Dower WJ, Cwirla SE: Recombinant Library Screening Methods.

10

WO 91/18980; Filed May 13, 1991, Published Dec. 12, 1991. Devlin JJ: Compositions and Methods for Identifying Biologically Active Molecules.

- 15 WO 91/19818; Filed June 20, 1990, Published Dec. 26, 1991. Dower WJ, Cwirla SE, Barrett RW: Peptide Library and Screening Systems.

WO 92/02536; Filed Aug. 1, 1991, Published Feb. 20, 1992. Gold L, Tuerk C: Systematic Polypeptide Evolution by Reverse Translation.

- 20 WO 92/03918; Filed Aug. 28, 1991, Published Mar. 19, 1992. Lonberg N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

WO 92/03918; Filed Aug. 28, 1991, Published Mar. 19, 1992. Lonberg N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

25

WO 92/05258; Filed Sept. 17, 1991, Published Apr. 2, 1992. Fincher GB: Gene Encoding Barley Enzyme.

WO 92/14843; Filed Feb. 21, 1992, Published Sept. 3, 1992. Toole JJ, Griffin LC, Bock LC, Latham JA, Muenchau DD, Krawczyk S: Aptamers Specific for Biomolecules and Method of Making.

- 5 WO 93/08278; Filed Oct. 15, 1992, Published Apr. 29, 1993. Schatz PJ, Cull MG, Miller JF, Stemmer WP: Peptide Library and Screening Method.

WO 93/12227; Filed Dec 17, 1992, Published June 24, 1993. Lonberg, N; Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

10

WO 93/12227; Filed Dec. 17, 1992, Published June 24, 1993. Lonberg N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

WO 94/25585; Filed Apr. 25, 1994, Published Nov. 10, 1994. Lonberg, N, Kay RM:

- 15 Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

WO 94/25585; Filed Apr. 25, 1994, Published Nov. 10, 1994. Lonberg N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

- Arslan T, Abraham AT, Hecht SM: Structurally altered substrates for DNA topoisomerase I. Effects of inclusion of a single 3'-deoxynucleotide within the scissile strand. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):515-30.
- Aupeix K, Toulme JJ: Binding of chemically-modified oligonucleotides to the double-stranded stem of an RNA hairpin. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1647-50.
- Bazzanini R, Manfredini S, Durini E, Groschel B, Cinatl J, Balzarini J, De Clercq E, Imbach JL, Perigaud C, Gosselin G: Prodrugs of Ara-CMP and Ara-AMP with a S-acyl-2-thioethyl (SATE) biolabile phosphate protecting group: synthesis and biological evaluation. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):971-2.
- Blackburn GM, Liu X, Rosler A, Brenner C: Two hydrolase resistant analogues of diadenosine 5',5'''-P1,P3-triphosphate for studies with Fhit, the human fragile histidine triad protein. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):301-8.
- Bridson PK, Lin X, Melman N, Ji XD, Jacobson KA: Synthesis and adenosine receptor affinity of 7-beta-D-ribofuranosylxanthine. *Nucleosides Nucleotides* 1998 Apr;17(4):759-68.

- Brodin P, Gottikh M, Auclair C, Mouscadet JF: Inhibition of HIV-1 integration by mono- & bi-functionalized triple helix forming oligonucleotides. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1717-8.
- Creighton TE: Proteins Structures and Molecular Principles. New York: W. H. Freeman and Co., 1984.
- De Clercq E: Carbocyclic adenosine analogues as S-adenosylhomocysteine hydrolase inhibitors and antiviral agents: recent advances. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):625-34.
- de Zwart M, Link R, von Frijtag Drabbe Kunzel JK, Cristalli G, Jacobson KA, Townsend-Nicholson A, IJzerman AP: A functional screening of adenosine analogues at the adenosine A2B receptor: a search for potent agonists. *Nucleosides Nucleotides* 1998 Jun;17(6):969-85.
- Egron D, Arzumanov AA, Dyatkina NB, Krayevsky A, Imbach JL, Aubertin AM, Gosselin G, Perigaud C: Synthesis, anti-HIV activity and stability studies of 3'-azido-2',3'-dideoxythymidine 5'-fluorophosphate. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):983-4
- Gianolio DA, McLaughlin LW: Synthesis and triplex forming properties of pyrimidine derivative containing extended functionality. *Nucleosides Nucleotides* 1999 Aug;18(8):1751-69.
- Gottikh MB, Volkov EM, Romanova EA, Oretskaya TS, Shabarova ZA: Synthesis of oligonucleotide-intercalator conjugates capable to inhibit HIV-1 DNA integration. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1645-6.
- Hotoda H, Koizumi M, Ohmine T, Furukawa H, Nishigaki T, Abe K, Kosaka T, Tsutsumi S, Sone J, Kaneko M: Biologically active oligodeoxyribonucleotides. 10: anti-HIV-1 activity and stability of modified hexanucleotides containing glycerol-skeleton. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):243-52.
- JP10113194; Filed 19971022, Published 19980506. Donnelly, JJ ;Dwarki, VJ ;Liu, MA ;Montgomery, DL ;Parker, S ;Shiver, JW ;Ulmer JB: Nucleic Acid Preparation.

- Kang SH, Sinhababu AK, Cho MJ: Synthesis and biological activity of bis(pivaloyloxymethyl) ester of 2'-azido-2'-deoxyuridine 5'-monophosphate. *Nucleosides Nucleotides* 1998 Jun;17(6):1089-98.
- Krayevsky A, Arzumanov A, Shirokova E, Dyatkina N, Victorova L, Jasko M, Alexandrova L: dNTP modified at triphosphate residues: substrate properties towards DNA polymerases and stability in human serum. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):681-93.
- Krayevsky AA, Dyatkina NB, Semizarov DG, Victorova LS, Shirokova EA, Theil F, Von Janta Lipinski MJ, Gosselin G, Imbach JL: Reasons and limits of substrate activity of modified L-dNTP in DNA biosynthesis. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):863-4.
- Kvasyuk EI, Mikhailopulo IA, Suhadolnik RJ, Henderson EE, Muto NF, Iacono KT, Homon J, Pfeleiderer W: Synthesis and biological activity of 2',5'-oligoadenylate trimers containing 5'-terminal 5'-amino-5'-deoxy- and 5'-amino-3',5'-dideoxyadenosine derivatives. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1483-4.
- Liu J, Skradis A, Kolar C, Kolath J, Anderson J, Lawson T, Talmadge J, Gmeiner WH: Increased cytotoxicity and decreased in vivo toxicity of FdUMP[10] relative to 5-FU. *Nucleosides Nucleotides* 1999 Aug;18(8):1789-802.
- Lutz MJ, Will DW, Breipohl G, Benner SA, Uhlmann E: Synthesis of a monocharged peptide nucleic acid (PNA) analog and its recognition as substrate by DNA polymerases. *Nucleosides Nucleotides* 1999 Mar;18(3):393-401.
- Monaco V, van de Wetering KI, Meeuwenoord NJ, van den Elst HA, Stuivenberg HR, Visse R, van der Kaaden JC, Moolenaar GF, Verhoeven EE, Goosen N, van der Marel GA, van Boom JH: Synthesis and biological evaluation of modified DNA fragments for the study of nucleotide excision repair in *E. coli*. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1339-41.

Morozova OV, Kolpashchikov DM, Ivanova TM, Godovikova TS: Synthesis of new photocross-linking 5-C-base-substituted UTP analogs and their application in highly selective affinity labelling of the tick-borne encephalitis virus RNA replicase proteins. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1513-4.

Nguyen-Ba N, Chan L, Quimpere M, Turcotte N, Lee N, Mitchell H, Bedard J: Design and SAR study of a novel class of nucleotide analogues as potent anti-HCMV agents. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):821-7.

- Pandolfi D, Rauzi F, Capobianco ML: Evaluation of different types of end-capping modifications on the stability of oligonucleotides toward 3'- and 5'-exonucleases. *Nucleosides Nucleotides* 1999 Sep;18(9):2051-69.
- Pankiewicz KW, Lesiak-Watanabe K: Novel mycophenolic adenine bis(phosphonate)s as potent anticancer agents and inducers of cells differentiation. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):927-32.
- Perrin DM, Garestier T, Helene C: Expanding the catalytic repertoire of nucleic acid catalysts: simultaneous incorporation of two modified deoxyribonucleoside triphosphates bearing ammonium and imidazolyl functionalities. *Nucleosides Nucleotides* 1999 Mar;18(3):377-91.
- Pfundheller HM, Koshkin AA, Olsen CE, Wengel J: Evaluation of oligonucleotides containing two novel 2'-O-methyl modified nucleotide monomers: a 3'-C-allyl and a 2'-O,3'-C-linked bicyclic derivative. *Nucleosides Nucleotides* 1999 Sep;18(9):2017-30.
- Ramasamy KS, Stoisavljevic V: Synthesis and biophysical studies of modified oligonucleotides containing acyclic amino alcohol nucleoside analogs. *Nucleosides Nucleotides* 1999 Aug;18(8):1845-61.
- Schinazi RF, Lesnikowski ZJ: Boron containing oligonucleotides. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):635-47.
- Secrist JA 3rd, Parker WB, Allan PW, Bennett LL Jr, Waud WR, Truss JW, Fowler AT, Montgomery JA, Ealick SE, Wells AH, Gillespie GY, Gadi VK, Sorscher EJ: Gene therapy of cancer: activation of nucleoside prodrugs with E. coli purine nucleoside phosphorylase. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):745-57.
- Shirokova EA, Shipitsin AV, Victorova LS, Dyatkina NB, Goryunova LE, Beabealashvili RS, Hamilton CJ, Roberts SM, Krayevsky AA: Modified nucleoside 5'-triphosphonates as a new type of antiviral agents. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):1027-8.

Srivastava TK, Friedhoff P, Pingoud A, Katti SB: Application of oligonucleoside methylphosphonates in the studies on phosphodiester hydrolysis by Serratia endonuclease. *Nucleosides Nucleotides* 1999 Sep;18(9):1945-60.

Stattel JM, Yanachkov I, Wright GE: Synthesis and biochemical study of N2-(p-n-butylphenyl)-2'-deoxyguanosine 5'-(alpha,beta-imido)triphosphate (BuPdGMPNHPP): a non-substrate inhibitor of B family DNA polymerases. *Nucleosides Nucleotides* 1998 Aug;17(8):1505-13.

Terato H, Morita H, Ohyama Y, Ide H: Novel modification of 5-formyluracil by cysteine derivatives in aqueous solution. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):131-41.

Tomikawa A, Seno M, Sato-Kiyotaki K, Ohtsuki C, Hirai T, Yamaguchi T, Kawaguchi T, Yoshida S, Saneyoshi M: Synthetic nucleosides and nucleotides. 40. Selective inhibition of eukaryotic DNA polymerase alpha by 9-(beta-D-arabinofuranosyl)-2-(p-n-butylanilino) adenine 5'-triphosphate (BuAaraATP) and its 2'-up azido analog: synthesis and enzymatic evaluations. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):487-501.

US5580859; Filed 19940318, Issued 19961203. Felgner, PL. ;Wolff, JA. ;Rhodes, GH. ;Malone, RW. ;Carson, DA.: Delivery of exogenous DNA sequences in a mammal.

US5589466; Filed 19950126, Issued 19961231. Felgner, PL. ;Wolff, JA. ;Rhodes, GH. ;Malone, RW. ;Carson, DA.: Induction of a protective immune response in a mammal by injecting a DNA sequence.

US5641665; Filed 19941128, Issued 19970624. Hobart, PM. ;Margalith, M ;Parker, SE. ;Khatibi, S: Plasmids suitable for IL-2 expression.

US5693622; Filed 19950607, Issued 19971202. Wolff, JA. ;Duke, DJ. ;Felgner, PL.: Expression of exogenous polynucleotide sequences cardiac muscle of a mammal.

US5703055; Filed 19940126, Issued 19971230. Felgner, PL. ;Wolff, JA ;Rhodes, GH. ;Malone, RW ;Carson, DA.: Generation of antibodies through lipid mediated DNA delivery.

US5846946; Filed 19960614, Issued 19981208. Huebner, RC. ;Norman, JA. ;Liang, X ;Carner, KR. ;Barbour, AG. ;Luke, CJ.: Compositions and methods for administering Borrelia DNA.

US5910488; Filed 19951201, Issued 19990608. Nabel, GJ. ;Nabel, EG. ;Lew, D ;Marquet, M: Plasmids suitable for gene therapy.

Victorova LS, Semizarov DG, Shirokova EA, Alexandrova LA, Arzumanov AA, Jasko MV, Krayevsky AA: Human DNA polymerases and retroviral reverse transcriptases: selectivity in respect to dNTPs modified at triphosphate residues. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):1031-2.

von Janta-Lipinski M, Gaertner K, Lehmann C, Scheer H, Schildt J, Matthes E: Protein and RNA of human telomerase as targets for modified oligonucleotides. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1719-20

WO9011092; Filed 19900321, A1 Published 19901004. Felgner, PL. ;Wolff, JA;Rhodes, GH. ;Malone, RW ;Carson, DA.: Expression Of Exogenous Polynucleotide Sequences In A Vertebrate.

WO9314778; Filed 19930121, A1 Published 19930805. Rhodes, GH. ;Dwarki, VJ.;Felgner, PL;Wang-Felgner, J;Manthorpe, M: Ex Vivo Gene Transfer.

WO9421797; Filed 19940314, A1 Published 19940929. Donnelly, JJ. ;Dwarki, VJ. ;Liu, MA. ;Montgomery, DL. ;Parker, SE. ;Shiver, JW. ;Ulmer, JB.: Nucleic Acid Pharmaceuticals.

WO9633736; Filed 19960426, A1 Published 19961031. Baruch DI; Pasloske BL; Howard, RJ: Malaria Peptides and Vaccines.

WO9735992; Filed 19970317, A1 Published 19971002. Hobart, PM. ;Liang, X:
Tetracycline Inducible/Repressible Systems.

WO9926663; Filed 19981120, A2 Published 19990603. Horton, H ;Parker,
S;Manthorpe, M; Felgner, P: Treatment Of Cancer Using Cytokine-Expressing
Polynucleotides And Compositions Therefor.

WO9941368; Filed 19990210, A2 Published 19990819. Punnonen J, Stemmer WP,
Whalen RG; Howard, R: Optimization of Immunomodulatory Properties of Genetic
Vaccines.

WO9941369; Filed 19990210, A2 Published 19990819. Punnonen J, Stemmer WP,
Whalen RG; Howard, R: Genetic Vaccine Vector Engineering.

WO9941383; Filed 19990210, A1 Published 19990819. Punnonen J, Bass, SH,
Whalen, RG, Howard, R, Stemmer, WP: Antigen Library Immunization.

WO9941402; Filed 19990210, A2 Published 19990819. Punnonen J, Stemmer, WP,
Howard R, Patten PA: Targeting of Genetic Vaccine Vectors.

3. CLAIMS

1. A method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising:

creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set;

wherein optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation and iterative manner;

whereby these directed evolution methods include the introduction of mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

2. The method of claim 1, wherein said optimized modulatory effect on an immune response is induced by a genetic vaccine vector.

3. A method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising:

screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically generated progeny polynucleotide that has, or

encodes a polypeptide that has, a modulatory effect on an immune response; wherein the optimized non-stochastically generated polynucleotide or the polypeptide encoded by the non-stochastically generated polynucleotide exhibits an enhanced ability to modulate an immune response compared to a parental polynucleotide from which the library was created.

4. The method of claim 3, wherein said optimized modulatory effect on an immune response is induced by a genetic vaccine vector.

5. A method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising:

- a) creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set; and
- b) screening the library to identify an optimized non-stochastically generated progeny polynucleotide that has, or encodes a polypeptide that has, a modulatory effect on an immune response induced by a genetic vaccine vector; wherein the optimized non-stochastically generated polynucleotide or the polypeptide encoded by the non-stochastically generated polynucleotide exhibits an enhanced ability to modulate an immune response compared to a parental polynucleotide from which the library was created;

whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation, and iterative manner;

whereby these directed evolution methods include the introduction of point mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

6. The method of claim 5, wherein said optimized modulatory effect on an immune response is induced by a genetic vaccine vector.

7. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide is incorporated into a genetic vaccine vector.

8. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide, or a polypeptide encoded by the optimized non-stochastically generated polynucleotide, is administered in conjunction with a genetic vaccine vector.

9. The method of any of claims 1-6, wherein the library of non-stochastically generated progeny polynucleotides is created by a process selected from the group consisting of gene reassembly, oligonucleotide-directed saturation mutagenesis, and any combination, permutation and iterative manner.

10. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide that has a modulatory effect on an immune response is obtained by:

a) non-stochastically reassembling at least two parental template polynucleotide, each of which is, or encodes a molecule that is, involved in modulating an immune response;

wherein the first and second parental templates differ from each other in two or more nucleotides, to produce a library of non-stochastically generated polynucleotides; and

b) screening the library to identify at least one optimized non-stochastically generated polynucleotide that exhibits, either by itself or through the encoded molecule, an enhanced ability to modulate an immune response in comparison to a parental polynucleotide from which the library was created.

11. The method of claim 10, wherein the method further comprises the steps of:

c) subjecting a working optimized non-stochastically generated polynucleotide to a further round of non-stochastic reassembly with at least one additional polynucleotide, which is the same or different from the first and second polynucleotides, to produce a further working library of recombinant polynucleotides;

d) screening the further working library to identify at least one further optimized non-stochastically generated polynucleotide that exhibits an enhanced ability to modulate an immune response in comparison to a parental polynucleotide from which the library was created; and

e) optionally repeating c) and d) as necessary, until a desirable further optimized non-stochastically generated polynucleotide that exhibits an enhanced ability to modulate an immune response than a form of the nucleic acid from which the library was created.

12. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a polypeptide that can interact with a cellular receptor involved in mediating an immune response; wherein the polypeptide acts as an agonist or antagonist of the receptor.

13. The method of claim 12, wherein the cellular receptor is a macrophage scavenger receptor.

14. The method of claim 12, wherein the cellular receptor is selected from the group consisting of a cytokine receptor and a chemokine receptor.

15. The method of claim 14, wherein the chemokine receptor is CCR6.

16. The method of claim 12, wherein the polypeptide mimics the activity of a natural ligand for the receptor but does not induce immune reactivity to said natural ligand.

17. The method of claim 12, wherein the library is screened by:

i) expressing the non-stochastically generated progeny polynucleotides so that the encoded polypeptides are produced as fusions with a protein displayed on the surface of a replicable genetic package;

- ii) contacting the replicable genetic packages with a plurality of cells that display the receptor; and
- iii) identifying cells that exhibit a modulation of an immune response mediated by the receptor.

18. The method of claim 17, wherein the replicable genetic package is selected from the group consisting of a bacteriophage, a cell, a spore, and a virus.

19. The method of claim 18, wherein the replicable genetic package is an M13 bacteriophage and the protein is encoded by geneIII or geneVIII.

20. The method of claim 12, which method further comprises introducing the optimized non-stochastically generated polynucleotide into a genetic vaccine vector and administering the vector to a mammal, wherein the peptide or polypeptide is expressed and acts as an agonist or antagonist of the receptor.

21. The method of claim 12, which method further comprises producing the polypeptide encoded by the optimized non-stochastically generated polynucleotide and introducing the polypeptide into a mammal in conjunction with a genetic vaccine vector.

22. The method of claim 12, wherein the optimized non-stochastically generated polynucleotide is inserted into an antigen-encoding nucleotide sequence of a genetic vaccine vector.

23. The method of claim 22, wherein the optimized non-stochastically generated polypeptide is introduced into a nucleotide sequence that encodes an M- loop of an HBsAg polypeptide.

24. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide comprises a nucleotide sequence rich in unmethylated CpG.

25. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a polypeptide that inhibits an allergic reaction.

26. The method of claim 25, wherein the polypeptide is selected from the group consisting of interferon- α , interferon- β , IL-10, IL-12, an antagonist of IL-4, an antagonist of IL-5, and an antagonist of IL-13.

27. The method of 1, wherein the optimized recombinant polynucleotide encodes an antagonist of IL-10.

28. The method of claim 27, wherein the antagonist of IL-10 is soluble or defective IL-10 receptor or IL-20/MDA-7.

29. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a co-stimulator.

30. The method of claim 29, wherein the co-stimulator is B7-1 (CD80) or B7-2 (CD86) and the screening step involves selecting variants with altered activity through CD28 or CTLA-4.

31. The method of claim 29, wherein the co-stimulator is CD1, CD40, CD154 (ligand for CD40) or CD150 (SLAM).

32. The method of claim 29, wherein the co-stimulator is a cytokine.

33. The method of claim 32, wherein the cytokine is selected from the group consisting of IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, GM-CSF, G-CSF, TNF-, IFN-, IFN-, and IL-20 (MDA-7).

34. The method of 33, wherein the library of non-stochastically generated polynucleotides is screened by testing the ability of cytokines encoded by the non-stochastically generated polynucleotides to activate cells which contain a receptor for the cytokine.

35. The method of claim 34, wherein the cells contain a heterologous nucleic acid that encodes the receptor for the cytokine.

36. The method of 33, wherein the cytokine is interleukin-12 and the screening is performed by: growing mammalian cells which contain the genetic vaccine vector in a culture medium; and detecting whether T cell proliferation or T cell differentiation is induced by contact with the culture medium.

37. The method of 33, wherein the cytokine is interferon- and the screening is performed by:

- i) expressing the non-stochastically generated polynucleotides so that the encoded polypeptides are produced as fusions with a protein displayed on the surface of a replicable genetic package;
- ii) contacting the replicable genetic packages with a plurality of B cells; and
- iii) identifying phage library members that are capable of inhibiting proliferation of the B cells.

38. The method of claim 33, wherein the immune response of interest is differentiation of T cells to T_H1 cells and the screening is performed by contacting a population of T cells with the cytokines encoded by the members of the library of recombinant polynucleotides and identifying library members that encode a cytokine that induces the T cells to produce IL-2 and interferon- .

39. The method of claim 32, wherein the cytokine encoded by the optimized non-stochastically generated polynucleotide exhibits reduced immunogenicity compared to a

cytokine encoded by a non-optimized polynucleotide, and the reduced immunogenicity is detected by introducing a cytokine encoded by the non-stochastically generated polynucleotide into a mammal and determining whether an immune response is induced against the cytokine.

40. The method of claim 29, wherein the co-stimulator is B7-1 (CD80) or B7-2 (CD86) and the cell is tested for ability to costimulate an immune response.

41. The method of any of claims 1-6, wherein the optimized recombinant polynucleotide encodes a cytokine antagonist.

42. The method of claim 41, wherein the cytokine antagonist is selected from the group consisting of a soluble cytokine receptor and a transmembrane cytokine receptor having a defective signal sequence.

43. The method of claim 41, wherein the cytokine antagonist is selected from the group consisting of IL-1 OR and IL-4R.

44. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a polypeptide capable of inducing a predominantly T_H1 immune response.

45. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a polypeptide capable of inducing a predominantly T_H2 immune response.

46. The method of any of claims 1-6, wherein said optimized modulatory effect on an immune response is a decrease in an unwanted modulatory effect on an immune response;

whereby application of the method can be used to generate a molecule having a decreased ability to elicit an immune response from a host recipient of said molecule, where said recipient can be a human or an animal host;

and whereby application of the method can thus be used to generate a molecule having decreased antigenicity with respect to at least one host recipient of said molecule.

47. The method of any of claims 1-6, wherein said optimized modulatory effect on an immune response is an increase in a desirable modulatory effect on an immune response;

whereby application of the method can be used to generate a molecule having an increased ability to elicit an immune response from a host recipient of said molecule, where said recipient can be a human or an animal host;

and whereby application of the method can thus be used to generate a molecule having increased antigenicity with respect to at least one host recipient of said molecule.

48. The method of any of claims 1-6, wherein said optimized modulatory effect on an immune response is both a decrease in a first unwanted modulatory effect on an immune response as well as an increase in a second desirable modulatory effect on an immune response;

whereby application of the method can be used to generate a molecule having both a decreased ability to elicit a first immune response from a first host recipient of said molecule as well as a an increased ability to elicit a second immune response from a second host recipient of said molecule;

whereby the first and the second recipient hosts can be the same or different;

whereby each of the first and the second recipient hosts can be a human or an animal host;

and whereby application of the method can thus be used to generate a molecule having both a first decreased antigenicity with respect to at least one host recipient of said molecule and a second decreased antigenicity with respect to at least one host recipient of said molecule.

49. The method of claim 48, wherein said first and said second modulatory effect on an immune response are evolved for respectively a first and a second module on the same multimodule vaccine vector;

whereby a module is exemplified by the following modules, as well as by a fragment derivative or analog thereof: an antigen coding sequence, a polyadenylation sequence, a sequence coding for a co-stimulatory molecule, a sequence coding for an inducible repressor or transactivator, a eukaryotic origin or replication, a prokaryotic origin of replication, a sequence coding for a prokaryotic marker, , and enhancer, a promoter, and operator, and an intron.

50. The method of any of claims 1-6, wherein the optimized modulatory effect on an immune response is comprised of an increase in the stability of the immunomodulatory (IM) polynucleotide or polypeptide encoded thereby;

whereby application of the method can be used to generate a molecule having an increased stability ex vivo, thus, for example, increasing shelf-life and/or ease of storage and/or length of time before expiration of activity upon storage;

and whereby application of the method can also be used to generate a molecule having an increased stability in vivo upon administration to a host recipient, thus, for example, increasing resistance to digestive acids and/or increasing stability in the circulation and/or any other method of elimination or destruction by the host recipient.

51. The method of any of claims 1-6, wherein the immunomodulatory (IM) polynucleotide or polypeptide encoded thereby; has an optimized modulatory effect on an immune response in a human host recipient;

whereby application of the method can thus be used to generate an optimized genetic vaccine for human recipients.

52. The method of any of claims 1-6, wherein the immunomodulatory (IM) polynucleotide or polypeptide encoded thereby; has an optimized modulatory effect on an immune response in an animal host recipient;

whereby application of the method can thus be used to generate an optimized genetic vaccine for animal recipients, including animals that are farmed or raised by man, animals that are not farmed or raised by man, domesticated animals, and non-domesticated animals.

53. A method for obtaining an optimized polynucleotide that encodes an accessory molecule that improves the transport or presentation of antigens by a cell, the method comprising:

- a) creating a library of non-stochastically generated polynucleotides by subjecting to optimization by non-stochastic directed evolution a parental polynucleotide set in which is encoded all or part of the accessory molecule; and
- b) screening the library to identify an optimized non-stochastically generated progeny polynucleotide that encodes a recombinant molecule that confers upon a cell an increased or decreased ability to transport or present an antigen on a surface of the cell compared to an accessory molecule encoded by template polynucleotides not subjected to the non-stochastic reassembly;

whereby application of the method can thus be used to generate an optimized molecule for human recipients &/or animal recipients, including animals that are farmed or raised by man, animals that are not farmed or raised by man, domesticated animals, and non-domesticated animals;

whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation, and iterative manner;

whereby these directed evolution methods include the introduction of point mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

54. The method of claim 53, wherein the screening involves:

- i) introducing the library of non-stochastically generated polynucleotides into a genetic vaccine vector that encodes an antigen to form a library of vectors; introducing the library of vectors into mammalian cells; and
- ii) identifying mammalian cells that exhibit increased or decreased immunogenicity to the antigen.

55. The method of claim 53, wherein the accessory molecule comprises a proteasome or a TAP polypeptide.

56. The method of claim 53, wherein the accessory molecule comprises a cytotoxic T-cell inducing sequence.

57. The method of claim 56, wherein the cytotoxic T-cell inducing sequence is obtained from a hepatitis B surface antigen.

58. The method of claim 53, wherein the accessory molecule comprises an immunogenic agonist sequence.

59. A method for obtaining an immunomodulatory polynucleotide that has, an optimized expression in a recombinant expression host, the method comprising:

creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set;

whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation and iterative manner;

whereby these directed evolution methods include the introduction of mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

60. A method for obtaining an immunomodulatory polynucleotide that has an optimized expression in a recombinant expression host, the method comprising:

screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically generated progeny polynucleotide that has an optimized expression in a recombinant expression host when compared to the expression of a parental polynucleotide from which the library was created.

61. A method for obtaining an immunomodulatory polynucleotide that has an optimized expression in a recombinant expression host, the method comprising:

a) creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set; and

b) screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically generated progeny polynucleotide that has an optimized expression in a recombinant expression host when compared to the expression of a parental polynucleotide from which the library was created;

whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation, and iterative manner;

whereby these directed evolution methods include the introduction of point mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

62. The method of any of claims 59-61, wherein the recombinant expression host is a prokaryote.

63. The method of any of claims 59-61, wherein the recombinant expression host is a eukaryote.

64. The method of claim 63, wherein the recombinant expression host is a plant.

65. The method of any of claims 64, wherein the recombinant expression host is a monocot.

66. The method of any of claims 64, wherein the recombinant expression host is a dicot.

67. The method of any of claims 1-6, 53, or 59-61, wherein creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set is comprised of subjecting the parental polynucleotide set to "gene site saturation mutagenesis" as described herein.

68. The method of any of claims 1-6, 53, or 59-61, wherein creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set is comprised of subjecting the parental polynucleotide set to "synthetic ligation polynucleotide reassembly" as described herein.

69. The method of any of claims 1-6, 53, or 59-61, wherein creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set is comprised of subjecting the parental polynucleotide set to both "gene site saturation mutagenesis" as described herein, and to "synthetic ligation polynucleotide reassembly" as described herein.

70. A method of producing a progeny polynucleotide set by subjecting a double-stranded circular parental polynucleotide molecule to mutagenesis, said method comprising the steps of:

a) annealing a first primer and a second primer to said parental polynucleotide molecule;

wherein said first primer is comprised of a first primer sequence that is complementary to a first annealment region of the parental polynucleotide molecule,

wherein said second primer is comprised of a second primer sequence that is complementary to a second annealment region of the parental polynucleotide molecule,

wherein said first annealment region and said second annealment region are non-overlapping and therefore staggered,

and wherein at least one of said first and second primers contains a non-stochastic mutagenic cassette with respect to the parental polynucleotide molecule; and

b) synthesizing by means of a polymerase-catalyzed amplification reaction a first progeny polynucleotide strand comprised of said first primer and a second progeny polynucleotide strand comprised of said second primer;

wherein the first progeny polynucleotide strand and the second progeny polynucleotide strand may form a double-stranded mutagenized circular polynucleotide product.

71. A method of producing a progeny polynucleotide set by subjecting a double-stranded circular parental polynucleotide molecule to mutagenesis, said method comprising the steps of:

a) annealing a first primer and a second primer to said parental polynucleotide molecule;

wherein said first primer is comprised of a first primer sequence that is complementary to a first annealment region of the parental polynucleotide molecule,
wherein said second primer is comprised of a second primer sequence that is complementary to a second annealment region of the parental polynucleotide molecule,
wherein said first annealment region and said second annealment region are non-overlapping and therefore staggered,
wherein at least one of said first and second primers contains a non-stochastic mutagenic cassette with respect to the parental polynucleotide molecule, and
wherein said non-stochastic mutagenic cassette contained in said at least one primer is degenerate in nature; and

b) synthesizing by means of a polymerase-catalyzed amplification reaction a first progeny polynucleotide strand comprised of said first primer and a second progeny polynucleotide strand comprised of said second primer;

wherein the first progeny polynucleotide strand and the second progeny polynucleotide strand may form a double-stranded mutagenized circular polynucleotide product;

whereby the generation of a degenerate progeny polynucleotide set may be achieved by applying said method.

72. A method for producing from a template polypeptide a set of progeny polypeptides in which a non-stochastic range of single amino acid substitutions is represented at each amino acid position, comprising the steps of:

a) subjecting a codon-containing template polynucleotide to polymerase-based amplification using a degenerate oligonucleotide for each codon to be mutagenized, wherein each of said degenerate oligonucleotides is comprised of a

first homologous sequence and a degenerate trinucleotide cassette, so as to generate a set of progeny polynucleotides; and

b) subjecting said set of progeny polynucleotides to clonal amplification such that polypeptides encoded by the progeny polynucleotides are expressed;

whereby, said method provides a means for generating a predetermined number of amino acids to be represented at each amino acid site along a parental polypeptide template, up to as many as all 20 amino acids at each of said amino acid sites.

73. The method of claim 72, wherein said degenerate oligonucleotide is comprised of a first homologous sequence, a degenerate trinucleotide cassette, and a second homologous sequence.

74. The method of claim 72, wherein said degenerate trinucleotide cassette is comprised of a first mononucleotide cassette selected from the group consisting of:

- a degenerate A/C mononucleotide cassette,
- a degenerate A/G mononucleotide cassette,
- a degenerate A/T mononucleotide cassette,
- a degenerate C/G mononucleotide cassette,
- a degenerate C/T mononucleotide cassette,
- a degenerate G/T mononucleotide cassette,
- a degenerate C/G/T mononucleotide cassette,
- a degenerate A/G/T mononucleotide cassette,
- a degenerate A/C/T mononucleotide cassette,
- a degenerate A/C/G mononucleotide cassette,
- and a degenerate N or A/C/G/T mononucleotide cassette;

and wherein said degenerate trinucleotide cassette is further comprised of a second and a third mononucleotide cassette, each selected from the group consisting of:

- a degenerate A/C mononucleotide cassette,
- a degenerate A/G mononucleotide cassette,
- a degenerate A/T mononucleotide cassette,
- a degenerate C/G mononucleotide cassette,
- a degenerate C/T mononucleotide cassette,
- a degenerate G/T mononucleotide cassette,
- a degenerate C/G/T mononucleotide cassette
- a degenerate A/G/T mononucleotide cassette,
- a degenerate A/C/T mononucleotide cassette,
- a degenerate A/C/G mononucleotide cassette,
- a degenerate N or A/C/G/T mononucleotide cassette,
- a non-degenerate A mononucleotide cassette,
- a non-degenerate C mononucleotide cassette,
- a non-degenerate G mononucleotide cassette,
- and a non-degenerate T mononucleotide cassette.

75. The method of claim 72, where said degenerate trinucleotide cassette is selected from the group consisting of:

- a degenerate N,N,N trinucleotide cassette,
- a degenerate N,N,G/T trinucleotide cassette,
- a degenerate N,N,G/C trinucleotide cassette,
- a degenerate N,N,A/C/G trinucleotide cassette,
- a degenerate N,N,A/G/T trinucleotide cassette,
- and a degenerate N,N,C/G/T trinucleotide cassette;

whereby, said method provides a means for generating all 20 amino acid changes at each amino acid site along a parental polypeptide template, because the degeneracy of the specified trinucleotide cassette sequences includes codons for all 20 amino acids.

76. The method of claim 72, wherein said degenerate oligonucleotide is comprised of a first homologous sequence and a plurality of trinucleotide cassettes;

whereby, said method provides a means for generating a progeny polypeptide having a plurality of concurrent single amino acid changes with respect to a parental polypeptide template.

77. The method of claim 76, wherein each of said degenerate trinucleotide cassettes is comprised of a first mononucleotide cassette selected from the group consisting of:

- a degenerate A/C mononucleotide cassette,
- a degenerate A/G mononucleotide cassette,
- a degenerate A/T mononucleotide cassette,
- a degenerate C/G mononucleotide cassette,
- a degenerate C/T mononucleotide cassette,
- a degenerate G/T mononucleotide cassette,
- a degenerate C/G/T mononucleotide cassette,
- a degenerate A/G/T mononucleotide cassette,
- a degenerate A/C/T mononucleotide cassette,
- a degenerate A/C/G mononucleotide cassette,

and a degenerate N or A/C/G/T mononucleotide cassette;

and wherein each of said degenerate trinucleotide cassettes is further comprised of a second and a third mononucleotide cassette, each selected from the group of consisting of:

- a degenerate A/C mononucleotide cassette,
- a degenerate A/G mononucleotide cassette,
- a degenerate A/T mononucleotide cassette,
- a degenerate C/G mononucleotide cassette,
- a degenerate C/T mononucleotide cassette,
- a degenerate G/T mononucleotide cassette,
- a degenerate C/G/T mononucleotide cassette
- a degenerate A/G/T mononucleotide cassette,
- a degenerate A/C/T mononucleotide cassette,
- a degenerate A/C/G mononucleotide cassette,
- a degenerate N or A/C/G/T mononucleotide cassette,
- a non-degenerate A mononucleotide cassette,
- a non-degenerate C mononucleotide cassette,
- a non-degenerate G mononucleotide cassette,
- and a non-degenerate T mononucleotide cassette.

78. The method of claim 76, where said degenerate trinucleotide cassette is selected from the group consisting of:

- a degenerate N,N,N trinucleotide cassette,
- a degenerate N,N,G/T trinucleotide cassette,
- a degenerate N,N,G/C trinucleotide cassette,
- a degenerate N,N,A/C/G trinucleotide cassette,

a degenerate N,N,A/G/T trinucleotide cassette,
and a degenerate N,N,C/G/T trinucleotide cassette;

whereby, said method provides a means for generating all 20 amino acid changes at each amino acid site along a parental polypeptide template, because the degeneracy of the specified trinucleotide cassette sequences includes codons for all 20 amino acids.

79. The method of claim 72, wherein said degenerate oligonucleotide is comprised of a first homologous sequence, and a plurality of trinucleotide cassettes, and a second homologous sequence.

80. A method for producing from a template polypeptide a set of progeny polypeptides in which a non-stochastic range of single amino acid substitutions is represented at each amino acid position, and for identifying desirable amino acid substitutions and combinations thereof among the progeny molecules, comprising the steps of:

- a) subjecting a codon-containing template polynucleotide to polymerase-based amplification using a degenerate oligonucleotide cassette for each codon to be mutagenized, wherein each of said degenerate oligonucleotides is comprised of a first homologous sequence and a degenerate trinucleotide cassette, so as to generate a set of progeny polynucleotides; and
- b) subjecting said set of progeny polynucleotides to clonal amplification such that polypeptides encoded by the progeny polynucleotides are expressed; and

c) subjecting said expressed progeny polypeptides to screening in order to compare them to the parental polynucleotide with respect to at least one molecular property of interest;

whereby, said method provides a means for generating a predetermined number of amino acids to be represented at each amino acid site along a parental polypeptide template, up to as many as all 20 amino acids at each of said amino acid sites; and

whereby, said method provides a means for identifying among said progeny polypeptides those that display a desirable change with respect to at least one molecular property when compared with its parental polypeptide.

81. The method of claim 80, wherein said degenerate trinucleotide cassette is comprised of a first nucleotide selected from the group consisting of:

- a degenerate A/C mononucleotide cassette,
- a degenerate A/G mononucleotide cassette,
- a degenerate A/T mononucleotide cassette,
- a degenerate C/G mononucleotide cassette,
- a degenerate C/T mononucleotide cassette,
- a degenerate G/T mononucleotide cassette,
- a degenerate C/G/T mononucleotide cassette,
- a degenerate A/G/T mononucleotide cassette,
- a degenerate A/C/T mononucleotide cassette,
- a degenerate A/C/G mononucleotide cassette,
- and a degenerate N or A/C/G/T mononucleotide cassette;

and wherein said degenerate trinucleotide cassette is further comprised of a second and a third mononucleotide cassette, each selected from the group consisting of:

a degenerate A/C mononucleotide cassette,
a degenerate A/G mononucleotide cassette,
a degenerate A/T mononucleotide cassette,
a degenerate C/G mononucleotide cassette,
a degenerate C/T mononucleotide cassette,
a degenerate G/T mononucleotide cassette,
a degenerate C/G/T mononucleotide cassette
a degenerate A/G/T mononucleotide cassette,
a degenerate A/C/T mononucleotide cassette,
a degenerate A/C/G mononucleotide cassette,
a degenerate N or A/C/G/T mononucleotide cassette,
a non-degenerate A mononucleotide cassette,
a non-degenerate C mononucleotide cassette,
a non-degenerate G mononucleotide cassette,
and a non-degenerate T mononucleotide cassette.

82. The method of claim 80, where said degenerate trinucleotide cassette is selected from the group consisting of:

a degenerate N,N,N trinucleotide cassette,
a degenerate N,N,G/T trinucleotide cassette,
a degenerate N,N,G/C trinucleotide cassette,
a degenerate N,N,A/C/G trinucleotide cassette,
a degenerate N,N,A/G/T trinucleotide cassette,
and a degenerate N,N,C/G/T trinucleotide cassette;

whereby, said method provides a means for generating all 20 amino acid changes at each amino acid site along a parental polypeptide template, because the

degeneracy of the specified trinucleotide cassette sequences includes codons for all 20 amino acids.

83. The method of claim 80, wherein said degenerate oligonucleotide is comprised of a first homologous sequence and a plurality of trinucleotide cassettes;

whereby, said method provides a means for generating a progeny polypeptide having a plurality of concurrent single amino acid changes with respect to a parental polypeptide template.

84. The method of claim 80, wherein each of said degenerate trinucleotide cassettes is comprised of a first mononucleotide cassette selected from the group consisting of:

- a degenerate A/C mononucleotide cassette,
- a degenerate A/G mononucleotide cassette,
- a degenerate A/T mononucleotide cassette,
- a degenerate C/G mononucleotide cassette,
- a degenerate C/T mononucleotide cassette,
- a degenerate G/T mononucleotide cassette,
- a degenerate C/G/T mononucleotide cassette,
- a degenerate A/G/T mononucleotide cassette,
- a degenerate A/C/T mononucleotide cassette,
- a degenerate A/C/G mononucleotide cassette,
- and a degenerate N or A/C/G/T mononucleotide cassette;

and wherein each of said degenerate trinucleotide cassettes is further comprised of a second and a third mononucleotide cassette, each selected from the group consisting of:

- a degenerate A/C mononucleotide cassette,
- a degenerate A/G mononucleotide cassette,
- a degenerate A/T mononucleotide cassette,
- a degenerate C/G mononucleotide cassette,
- a degenerate C/T mononucleotide cassette,
- a degenerate G/T mononucleotide cassette,
- a degenerate C/G/T mononucleotide cassette
- a degenerate A/G/T mononucleotide cassette,
- a degenerate A/C/T mononucleotide cassette,
- a degenerate A/C/G mononucleotide cassette,
- a degenerate N or A/C/G/T mononucleotide cassette,
- a non-degenerate A mononucleotide cassette,
- a non-degenerate C mononucleotide cassette,
- a non-degenerate G mononucleotide cassette,
- and a non-degenerate T mononucleotide cassette.

85. The method of claim 80, where said degenerate trinucleotide cassette is selected from the group consisting of:

- a degenerate N,N,N trinucleotide cassette,
- a degenerate N,N,G/T trinucleotide cassette,
- a degenerate N,N,G/C trinucleotide cassette,
- a degenerate N,N,A/C/G trinucleotide cassette,
- a degenerate N,N,A/G/T trinucleotide cassette,
- and a degenerate N,N,C/G/T trinucleotide cassette;

whereby, said method provides a means for generating all 20 amino acid changes at each amino acid site along a parental polypeptide template, because the degeneracy of the specified trinucleotide cassette sequences includes codons for all 20 amino acids.

Exo III Generated Structures

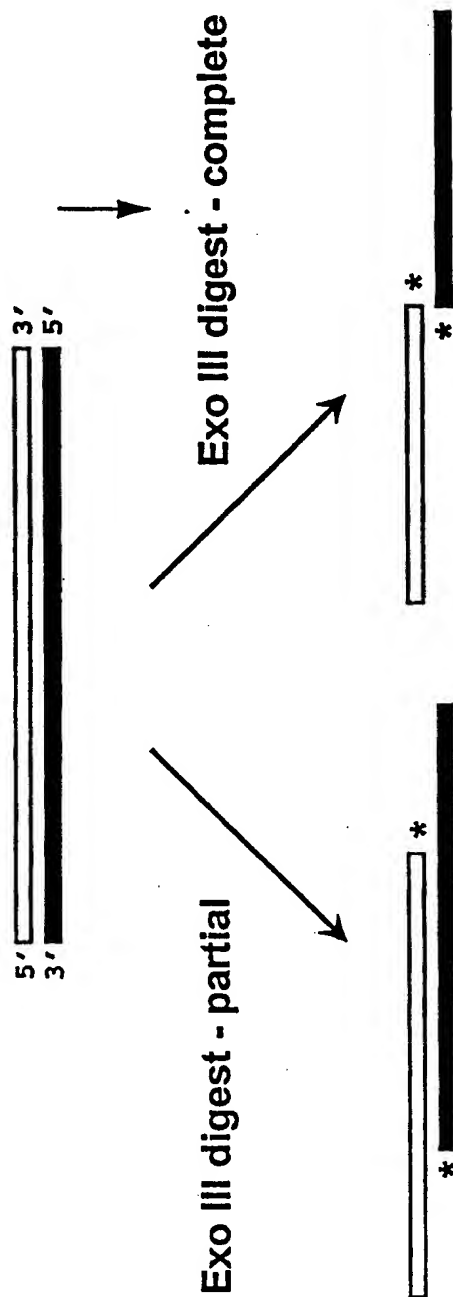


Figure 1

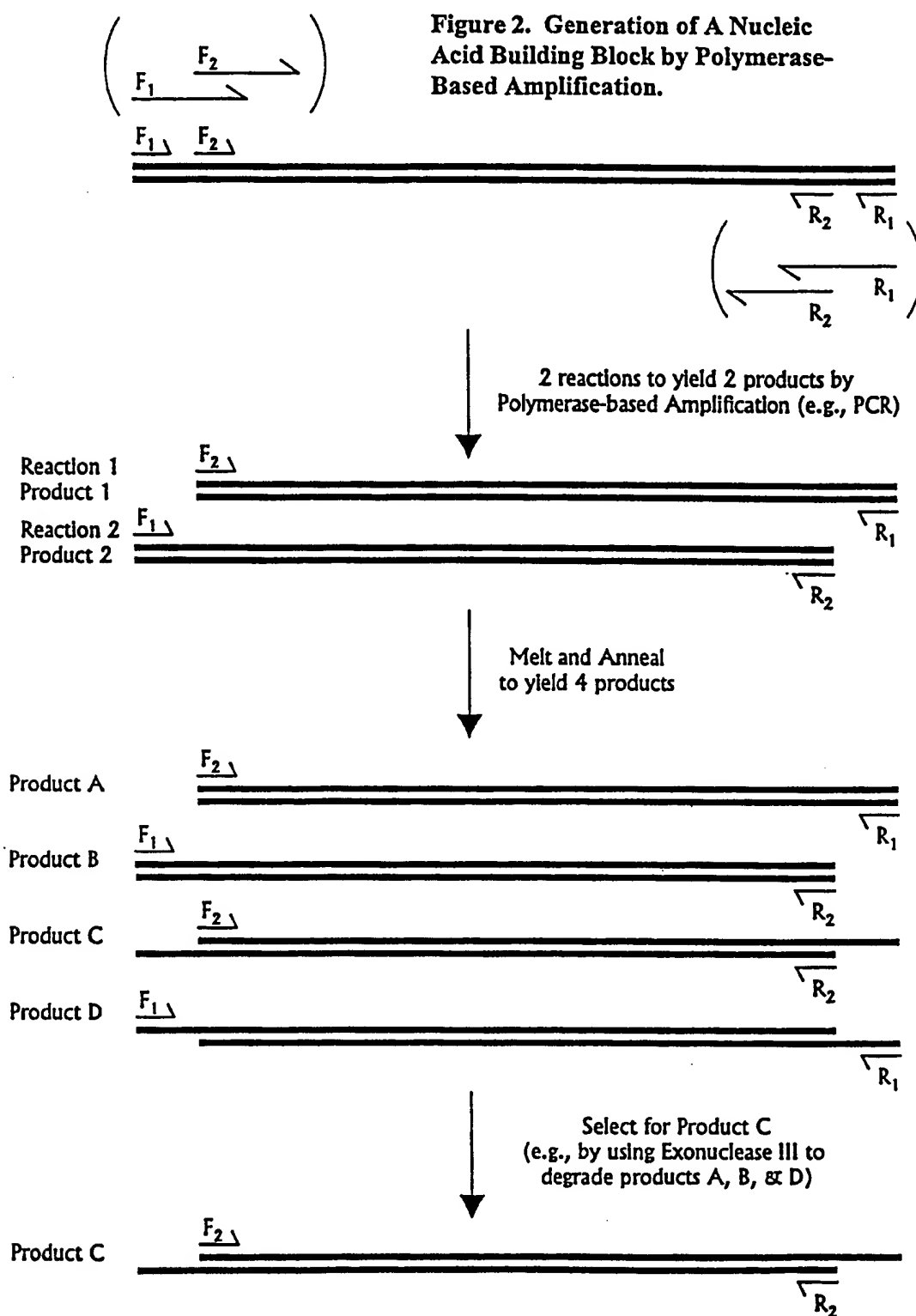
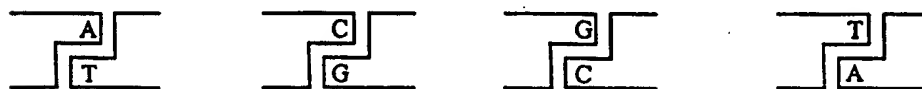


FIGURE 3. Unique Overhangs And Unique Couplings.

The number of unique overhangs of each size (e.g. the total number of unique overhangs composed of 1 or 2 or 3, etc. nucleotides) exceeds the number of unique couplings that can result from the use of all the unique overhangs of that size. For example, the total number of unique couplings that can be made using all the 8 unique single-nucleotide 3' overhangs and single-nucleotide 5' overhangs is 4.

PANEL A. 4 unique single-nucleotide 3' overhangs are possible (i.e., A, C, G, & T). For each of these there is a complementary 3' overhang with which it can pair (i.e., T, G, C, & A, respectively), as shown.



PANEL B. However, the number of unique single-nucleotide 3' overhangs is greater than the number of unique couplings. Thus, only 2 intrinsically unique couplings exist using single-nucleotide 3' overhangs as shown.



PANEL C. Likewise, 4 unique-single nucleotide 5' overhangs are possible (i.e., A, C, G, & T). For each of these there is a complementary 5' overhang with which it can pair (i.e., T, G, C, & A, respectively), as shown.



PANEL D. However, the number of unique single-nucleotide 5' overhangs is greater than the number of unique couplings. Thus, only 2 intrinsically unique couplings exist using single-nucleotide 5' overhangs as shown.

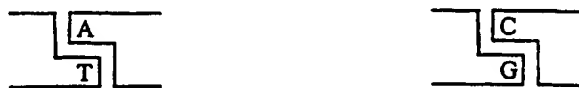
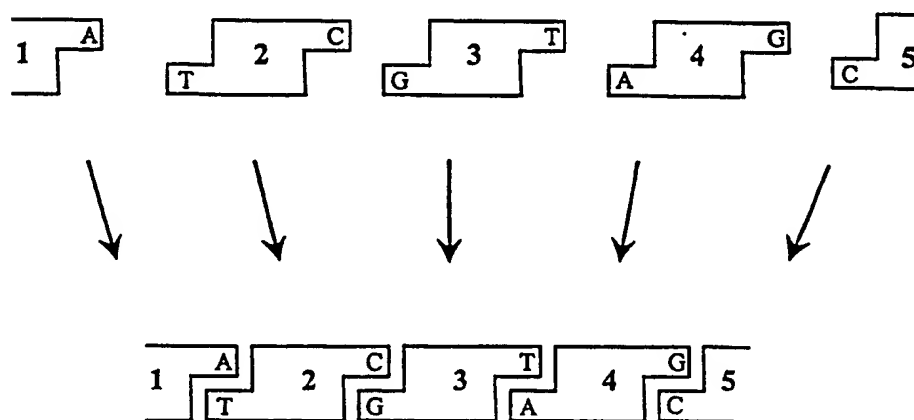


FIGURE 4. Unique Overall Assembly Order Achieved by Sequentially Coupling the Building Blocks

Awareness of the degeneracy (between the number of unique overhangs and the number of unique couplings) is important in order to avoid the production of degeneracy in the overall assembly order of the finalized nucleic acid. However, a unique overall assembly order can also be achieved - despite the use of non-unique couplings - by using building blocks having distinct combinations of couplings, and/or by stepping the assembly of the building blocks in a deliberately chosen sequence.

PANEL A. For example, one could attempt to assemble the following nucleic acid product using the 5 nucleic acid building blocks as shown.



PANEL B. However, degeneracy in the overall assembly order of the 5 nucleic acid building blocks would be present if the assembly process were carried out in one step. For example, building block #2 and building block #3 could both couple to building block #1 as shown.



FIGURE 4 cont.

PANEL C. However, a unique overall assembly order could be achieved by sequentially coupling the building blocks in 2 steps (rather than all at once) as shown.

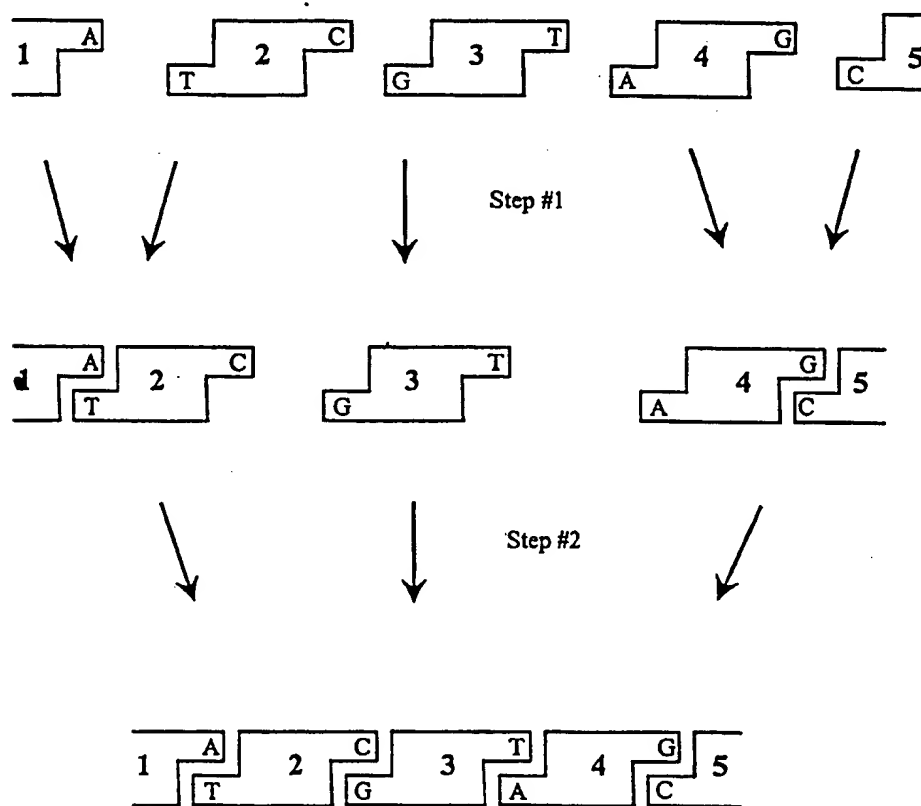


Figure 5. Unique Couplings Available Using a Two-Nucleotide 3' Overhang.

16 unique 3' overhangs can be formed using two-nucleotides. However, use of these 16 unique overhangs allows for the formation of only 6 unique couplings. Another 6 unique couplings are provided by the use 5' overhangs formed using two-nucleotides. Thus, a total of 12 unique couplings are provided by the combined use of 3' and 5' two-nucleotide overhangs. "Twin" couplings are marked in the same shading.

		TOP STRAND 2 ND Overhanging Nucleotide (counting from 5' to 3')					
		A	C	G	T		
TOP STRAND 1 ST Overhanging Nucleotide (counting from 5' to 3')	A	 	 	 	PALINDROMIC 	BOTTOM STRAND 2 ND Overhanging Nucleotide (counting from 5' to 3')	T
	C	 	 	PALINDROMIC 	 		G
	G	 	PALINDROMIC 	 	 		C
	T	PALINDROMIC 	 	 	 		A
		T	G	C	A		
		BOTTOM STRAND 1 ST Overhanging Nucleotide (counting from 5' to 3')					

Figure 6. Generation of an Exhaustive Set of Chimeric Combinations by Synthetic Ligation Reassembly.

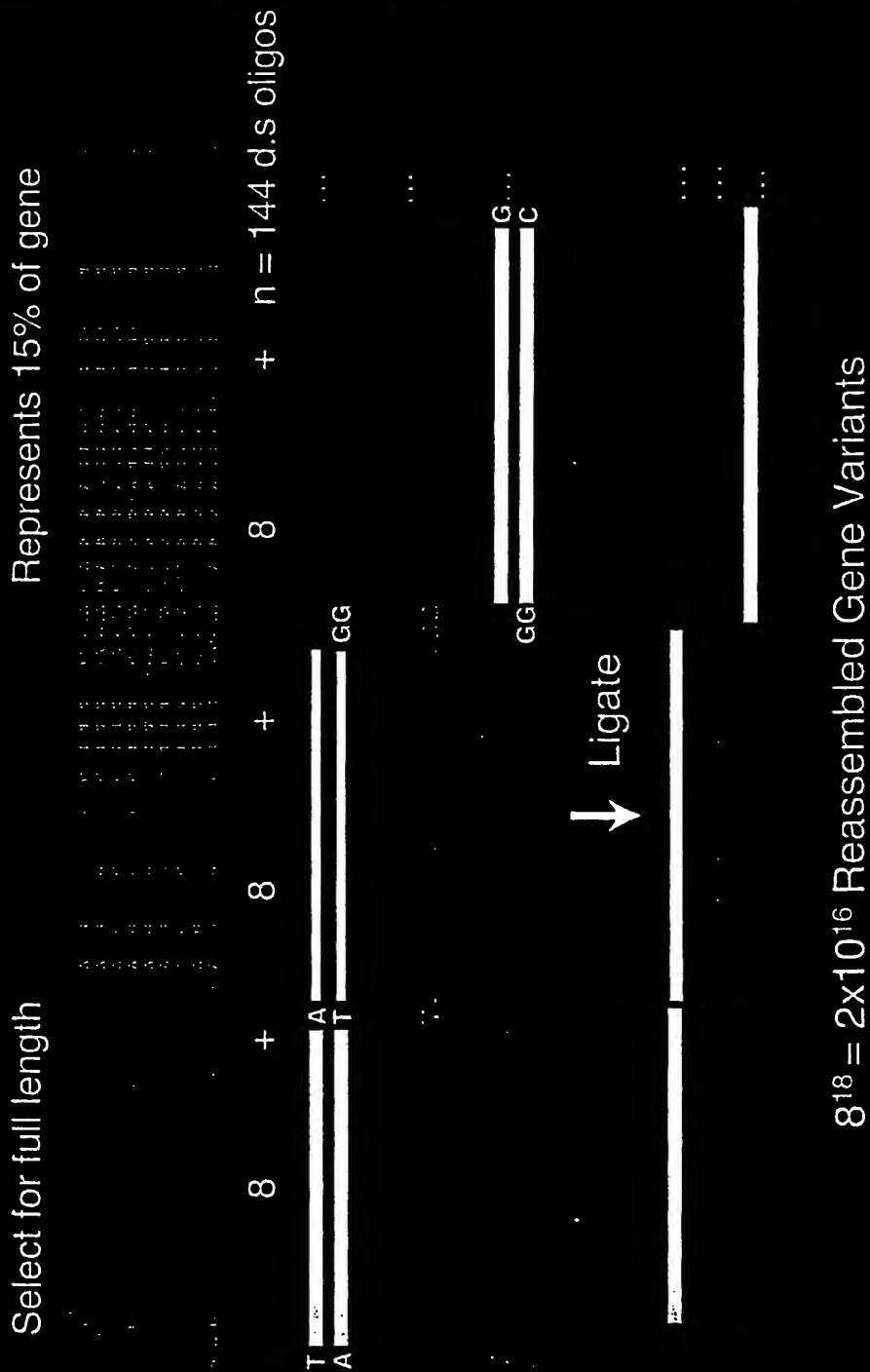


Figure 7. Synthetic genes from oligos.

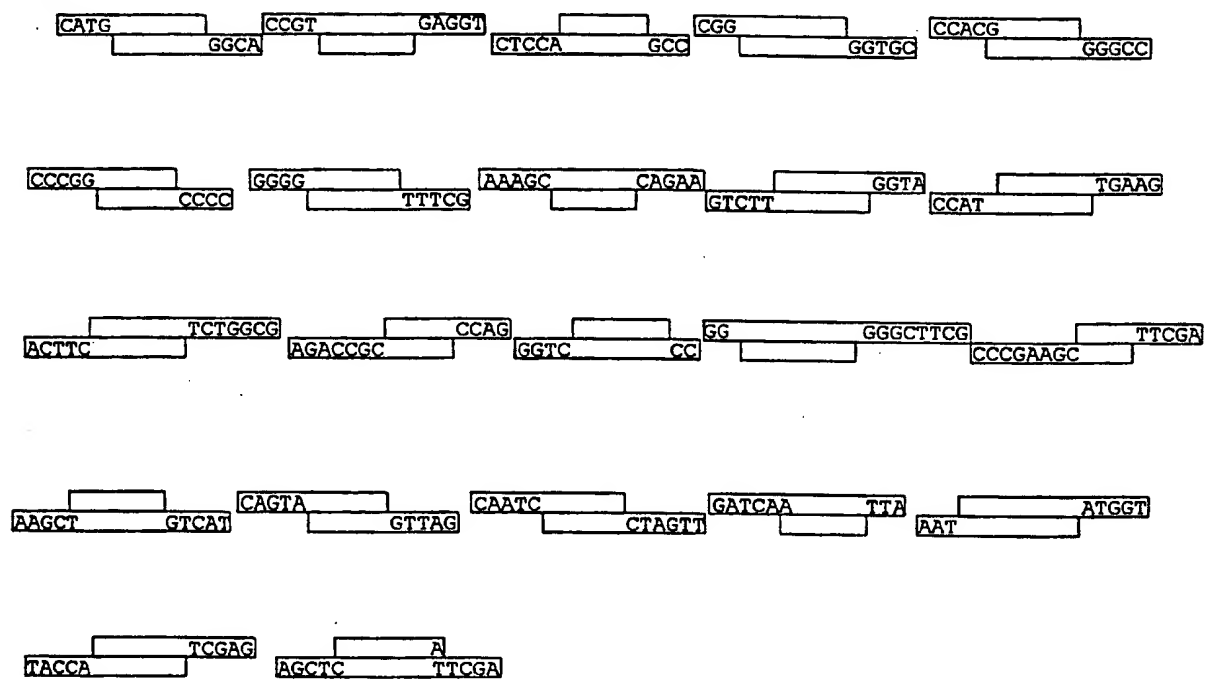
	NcoI					CCGT
150am13_00	c	ATGATGCACG	GCGATATTTTC	ATCGAGCAAT	GACACGGTCG	GCGTTGCCGT
150AM7_001	c	ATGCATCACG	GCGACATTTTC	ATCGAGCAAT	GACACGGTCG	GCGTTGCCGT
431am7_002	c	ATGAGACACG	GAGATATCTC	CAGCAGCAAC	GATTGCGTGG	GCGTGGCCGT
					GAG GT	
150am13_00		CGTGAACTAC	AAGATGCCTC	GCCTTCATAC	CAAGGCCGAG	GTTTTAGCGA
150AM7_001		CGTGAACTAC	AAGATGCCGC	GCCTTCACAC	CAAGGCCGAG	GTGCTGGCCA
431am7_002		CGTGAACTAC	AAGATGCCGC	GGCTGCATAC	CCGCGCCGAG	GTGATGGAGA
					CGG	
150am13_00		ACGCCAGAAA	GATCGGCGAG	ATGATCGTCG	GCATGAAGAC	CGGCCTGCCC
150AM7_001		ACTGCCGCAA	GATCGCCGAC	ATGCTGGTCG	GCATGAAGAG	CGGCCTGCCC
431am7_002		ACGCCCGCAA	GATCGCCGAC	ATGGTCGTGG	GCATGAAGCG	CGGCCTGCCC
					CCACG	
150am13_00		GGAATGGATC	TGGTGATCTT	CCCGGAATAT	TCGACCCACG	GCATCATGTA
150AM7_001		GGAATGGATC	TGGTGATCTT	CCCGGAATAT	TCCACCCACG	GCATCATGTA
431am7_002		GGCATGGACC	TGGTCATCTT	CCCCGAGTAC	TCCACCCACG	GCATCATGTA
					CCC GG	
150am13_00		CGACTCCAAG	GAAATGTACG	ATACCGCGTC	CGTCGTGCCG	GGCGAGGAGA
150AM7_001		CGACTCCAAG	GAGATGTACG	ACACGGCGTC	GACGGTGCCG	GGTGAAGAGA
431am7_002		CGACGCCAAG	GAAATGTACG	AAACCGCTTC	GGCCATGCCG	GGCGAAGAGA
					G GGG	
150am13_00		CCGAGATTTT	TGCCGAAGCC	TGCCGCAAGG	CGAAGTCTGG	GGGCGTGTTC
150AM7_001		CCGAGATTTT	CGCCGAGGCC	TGCCGCAAGG	CCAAGTCTGG	GGGCGTGTTC
431am7_002		CTGCTGTGTT	CGCCGACGCC	TGCCGCAAGG	CCAACGTATG	GGGCGTGTTC
					AAAG C	
150am13_00		TCGCTCACCG	GCGAACGTCA	CGAGGAACAT	CCGAAGCAAGG	CGCCCTACAA
150AM7_001		TCGCTGACCG	GCGAGCGCCA	CGAGGAGCAT	CCCAATAAAG	CGCCGTACAA
431am7_002		TCGCTGACCG	GCGAGCGCCA	CGAAGAGCAC	CCGAAGCAAGG	CGCCGTACAA
					CAG AA	
150am13_00		CACGCTGATC	CTGATGAACG	ACAAGGGCGA	GGTGGTCAG	AAATACCGCA
150AM7_001		CACCTGATC	CTGATGAACG	ACAAGGGTGA	AGTCGTCAG	AAATATCGCA
431am7_002		CACGCTCATC	CTGATGAACA	ACAAGGGCGA	GATCGTCAG	AAGTACCGCA
					GGTA	
150am13_00		AGATCATGCC	GTGGGTTCGG	ATCGAGGGCT	GGTATCCCCG	CAACTGCACC
150AM7_001		AGATCATGCC	GTGGGTGCCG	ATCGAAGGCT	GGTATCCCCG	CAACTGCACC
431am7_002		AGATCATGCC	CTGGGTGCCG	ATCGAAGGCT	GGTATCCCCG	CGATTGCACC
					TGAAG	
150am13_00		TACGTCTCCG	ACGGGCCGAA	GGGCATGAAG	GTTTCGCTGA	TCATCTGCGA
150AM7_001		TACGTCTCCG	AAGGCCCGAA	GGGCATGAAG	ATGTCGCTGA	TCATCTGCGA
431am7_002		TATGTGTCCG	AAGGCCCGAA	GGGACTGAAG	ATCAGCCTCA	TCATCTGCGA
					TCTGGCG	
150am13_00		TGACGGCAAC	TATCCGGAAA	TCTGGCGCGA	CTGCGCCATG	AAGGGCGCCG
150AM7_001		CGACGGCAAC	TATCCGGAAA	TCTGGCGTGA	CTGCGCGATG	AAGGGCGCCG
431am7_002		CGACGGCAAT	TATCCCGAGA	TCTGGCGCGA	TTGCGCCATG	CGGGCGCCG

Figure 7 cont.

	CCAG				
150am13_00	AGCTGATCGT	GCGCTGCCAG	GGCTACATGT	ATCCGGCCAA	GGACCAGCAG
150AM7_001	AACTGATCAT	CCGCTGCCAG	GGCTACATGT	ATCCCGCCAA	GGATCAGCAG
431am7_002	AGCTGATCGT	GCGTTGCCAG	GGATACATGT	ACCCGGCCAA	GGACCAGCAG
	GC				
150am13_00	GTCATCATGG	CGAAGGCAT	GGCGTGGGCG	AATAATTGTT	ACGTCGCGGT
150AM7_001	GTGCTGATGG	CGAAAGCAAT	GGCCTGGGCC	AACAACGTTT	ATGTCGCGGT
431am7_002	GTCATGGTGT	CCAAGGCAT	GGCGTGGATG	AACAACGTCT	ACGTGGCGGT
	GGGCTTCG				
150am13_00	TTCCAATGCC	GCGGGCTTCG	ATGGCGTCTA	TTCGTATTTC	GGCCACTCGG
150AM7_001	CGCCAATGCC	TGGGGCTTCG	ACGGCGTCTA	CTCGTATTTC	GGCCATTTCG
431am7_002	GGCCAATGCC	GCGGGCTTCG	ACGGCGTGTA	TTCCTACTTC	GGCCATTTCG
	TTCGA				
150am13_00	CGATCATCGG	CTTCGATGGC	CGCACGCTCG	GCGAATGCGG	CGAGGAAGAA
150AM7_001	CGATCATCGG	CTTCGACGGC	CGTACCCTCG	GCGAATGCGG	CGAGGAGGAT
431am7_002	CCATCATCGG	CTTCGACGGC	CGCACGCTGG	GCGAATGCGG	TGAAGAAGAC
	C AGTA				
150am13_00	TACGGCATCC	AGTATGCCCA	GCTTTCGAAG	ATGCTGATCC	GCGACGCCCC
150AM7_001	TATGGCATCC	AGTATGCCCG	CATCTCCAAG	TCGCTGATCC	GCGACGCGCG
431am7_002	ATGGGCGTGC	AGTACGCCGA	GCTCTCCACC	AGCCTGATCC	GCGACGCGCG
	CAATC				
150am13_00	CCGCACCGGA	CAATCGGAAA	ACCATCTCTT	CAAGCTGGTG	CATCGTGGCT
150AM7_001	CCGCACCGGC	CAATCGGAAA	ACCATCTCTT	CAAGCTGGTG	CACCGTGGCT
431am7_002	CAAGAACATG	CAGTCGCGAA	ACCACTTGTT	CAAGCTGGTG	CACCGCGGCT
	GATCAA				
150am13_00	ACACCGGGTT	GATCAATTC	GGCGAGGGCG	ACCGCGGTCT	CGCGGCTTGT
150AM7_001	ACACCGGCAT	GATCAATTC	GGCGAGGGCG	ACCGCGGTGT	CGCGGCTTGC
431am7_002	ACACCGGCAA	GATCAATTC	GGCGAAGAGG	CCACCGGCGT	CGCGGCATGC
	TTA				
150am13_00	CCCTTAAGAGT	TCTACAACAA	ATGGATCGCC	GATCCGGAAG	GCACCCGCGA
150AM7_001	CCGTAAGATT	TCTATTCGAA	ATGGATCGCC	GATCCCGAGG	GTACACGCGA
431am7_002	CCGTAAGACT	TCTACGCCAA	CTGGATCAAC	GATCCGGAGG	GCACGCGCAA
	ATGGT				
150am13_00	AATGGTCGAG	TCCTTTACCC	GGCCGACGGT	GGGAACCGAT	GAAGCGCCCA
150AM7_001	GATGGTGGAA	TCCTTCACGC	GTCCGACGGT	GGGTGTGGAG	GAATGCCCGA
431am7_002	GATGGTCGAA	TCCTTCACCC	GGTCCACCGT	GGGCACGCCG	GAGTGCCCA
	TCGAG				
150am13_00	TCCGAAGGCAT	CCCGAACAAAG	GTCGCGGTGC	ACCGCTGA	aagct
150AM7_001	TCCGAAGGCAT	TCCGAACAAAG	GCCACCACGC	ACCGCTGA	aagct
431am7_002	TGGACGGCAT	CCCCAACGAG	GACGCCAAGC	ACCGCTAG	aagct
					HindIII

Figure 8. Nucleic acid building blocks for synthetic ligation gene reassembly.

NcoI



HindIII

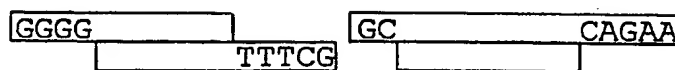
Figure 9. Addition of Introns by Synthetic Ligation Reassembly.

NcoI



Figure 10. Ligation Reassembly Using Fewer Than All The Nucleotides Of An Overhang.

Gap Ligation



Ligation of one strand only;
gap in second strand can be repaired in vivo

Figure 11. Avoidance of unwanted self-ligation in palindromic couplings.

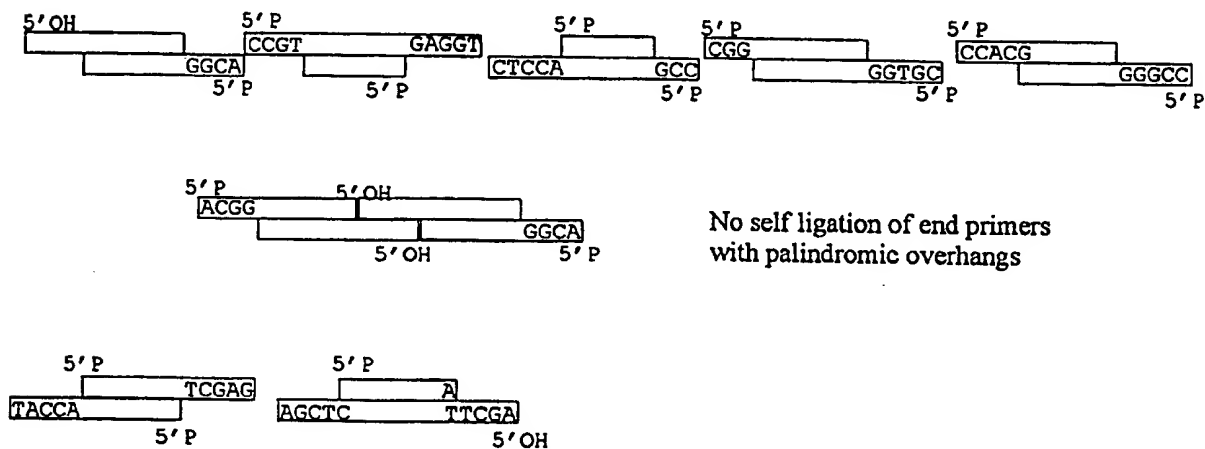
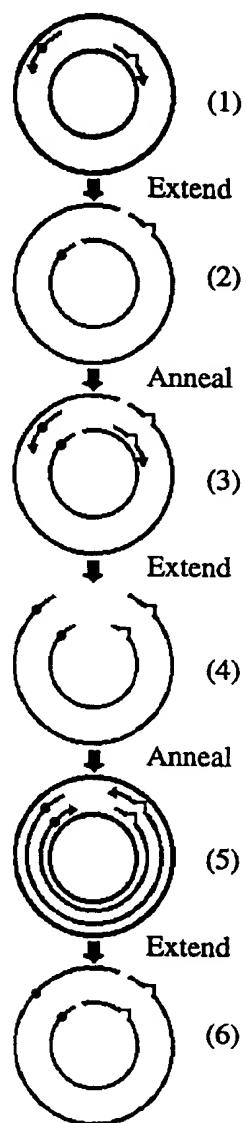


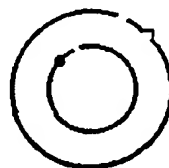
Figure 12
Site-Directed Mutagenesis by Polymerase-based Extension

Panel A

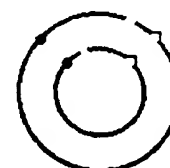


Panel B

Amplification products are comprised of the following molecular structures:



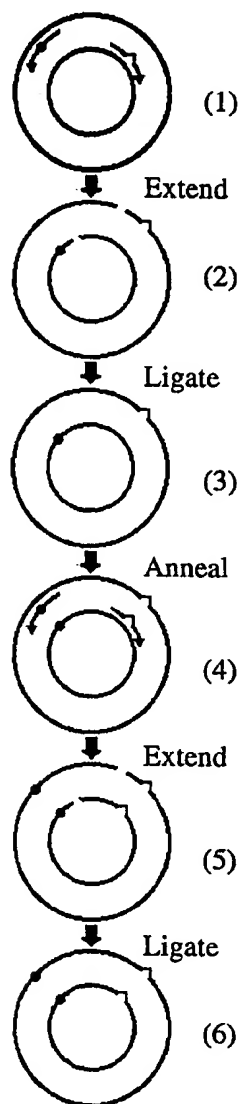
Molecule (A)



Molecule (B)

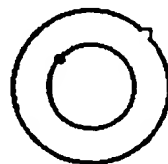
Figure 13
Site-Directed Mutagenesis By Polymerase-based Extension
and Ligase-based Ligation

Panel A

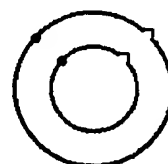


Panel B

Amplification products are comprised of the following molecular structures:



Molecule (A)



Molecule (B)

Figure 14

Strategy for obtaining and using nucleic acid binding proteins that facilitate entry of genetic vaccines.

Evolution in M13 Format

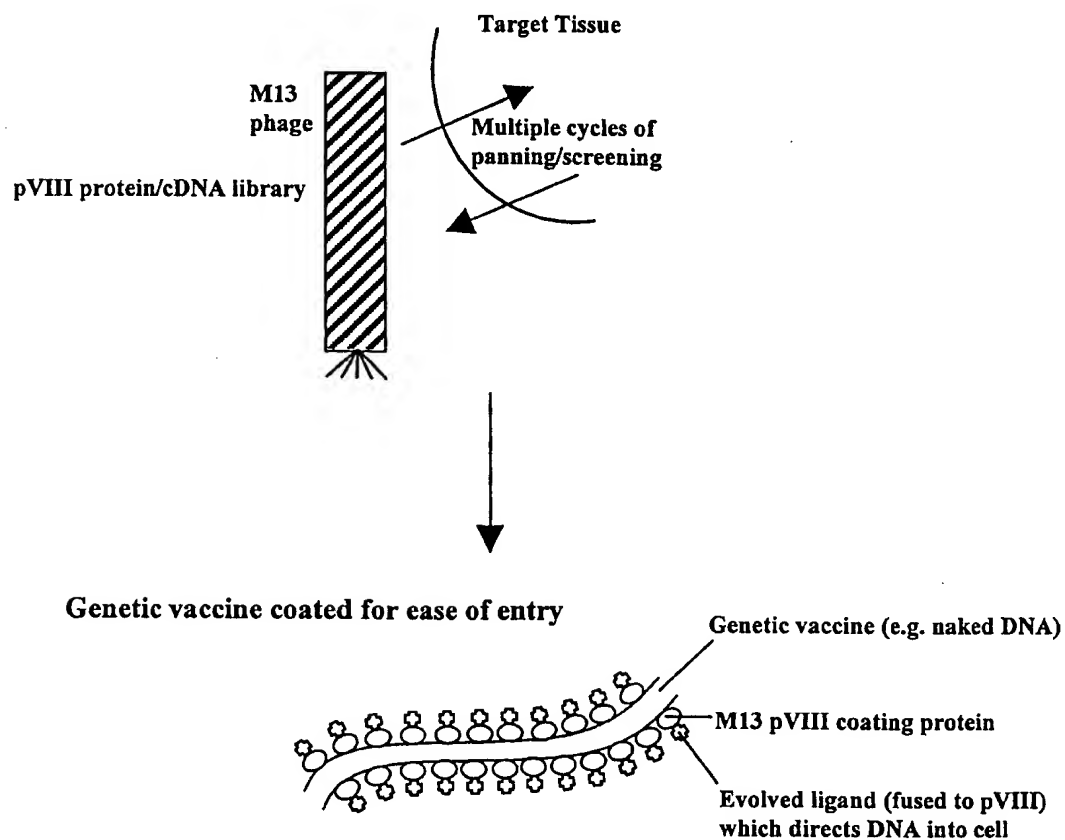


Figure 15

A schematic representation of a method for evolving a chimeric, multivalent antigen that has immunogenic regions from multiple antigens.

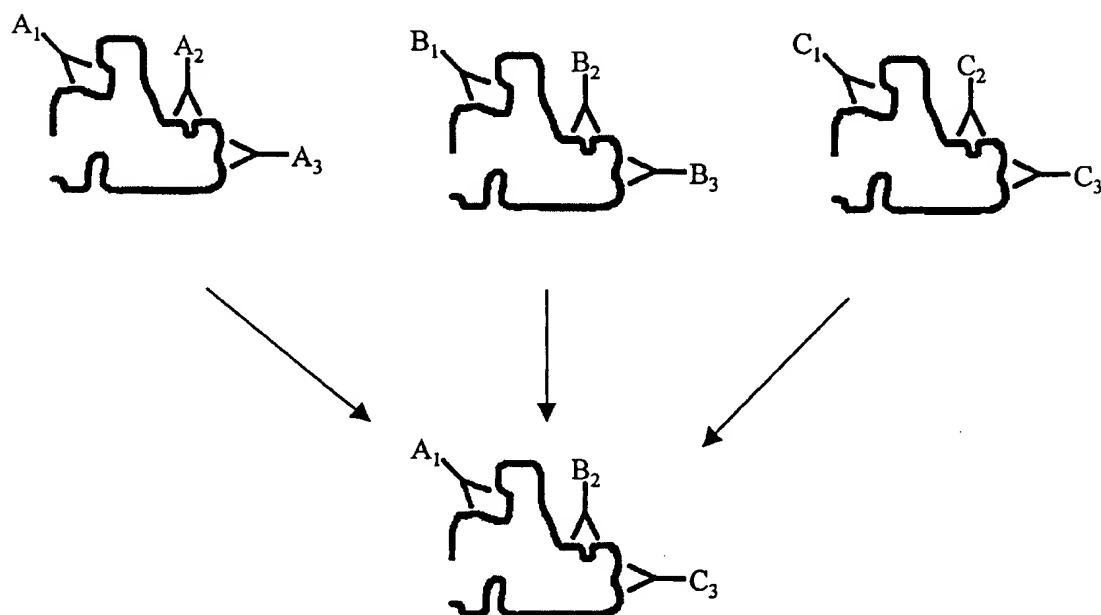


Figure 16

Method for Obtaining Non-Stochastically Generated Polypeptides that can induce a Broad-Spectrum Immune Response.

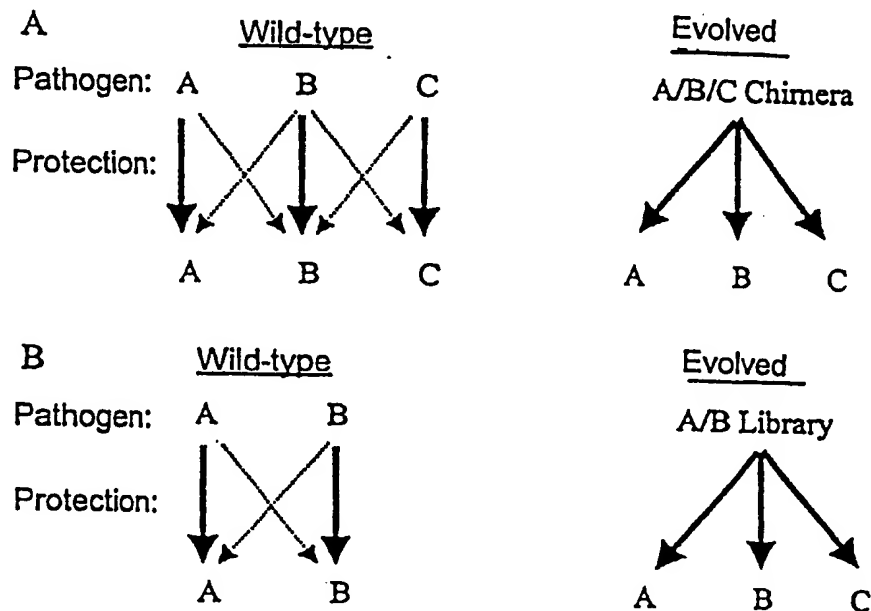


Figure 17

Possible factors for determining whether a particular polynucleotide encodes an immunogenic polypeptide having a desired property.

DNA - mRNA - Protein - Processing Manufacturing Antigen quality	Transcription	+ - + + - + + - + - + - + + + + - +	Pool of related antigen genes
	mRNA Stability	+ + - + + - - + + - + + - + + - + -	
	Translation	+ + + + - + - + + - + + - + - - + + + +	
	Codon Usage	- + - - + + - - + + - - + + + + - + + -	
	Protein Folding	+ + - + - + + - - + + - + + - + - + + +	
	Protein Stability	- + + + - + + - - + + - + + + + - + + +	
	B-cell Epitopes	+ + - + - + + - + + - + + - + - + + - +	
	T-cell Epitopes	- + - - + + + - + + - - + + + + - + + -	
	Ag Processing	+ + + - - + + - - + - - + + + - - + + + - +	
	TAP Binding	- + - - + + + - - + + - - - + + + + - +	
	HSP Binding	+ + + + - + + - - + + - + + - + - + + + +	
	Pept-ER Transport	+ - + + - + + - + - - + + - + + + + + -	
	Pept-MHC Binding	+ + - + + - - + + - + + - + - - - + + +	
	↓ Screen		
+ + + + + + + + + + + + + + - + + + + + +			

Figure 18

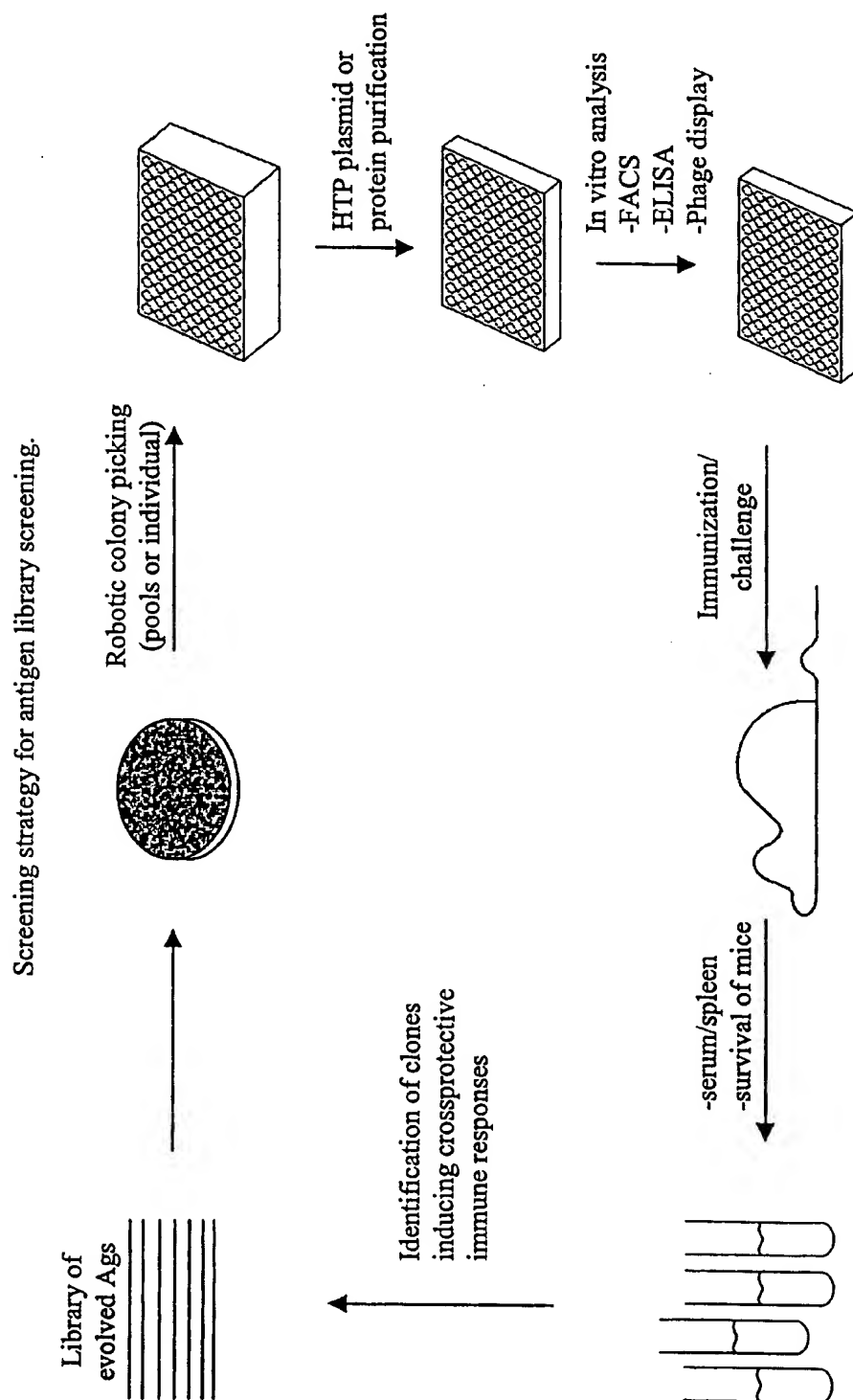


Figure 19

Strategy for pooling and deconvolution as used in antigen library screening

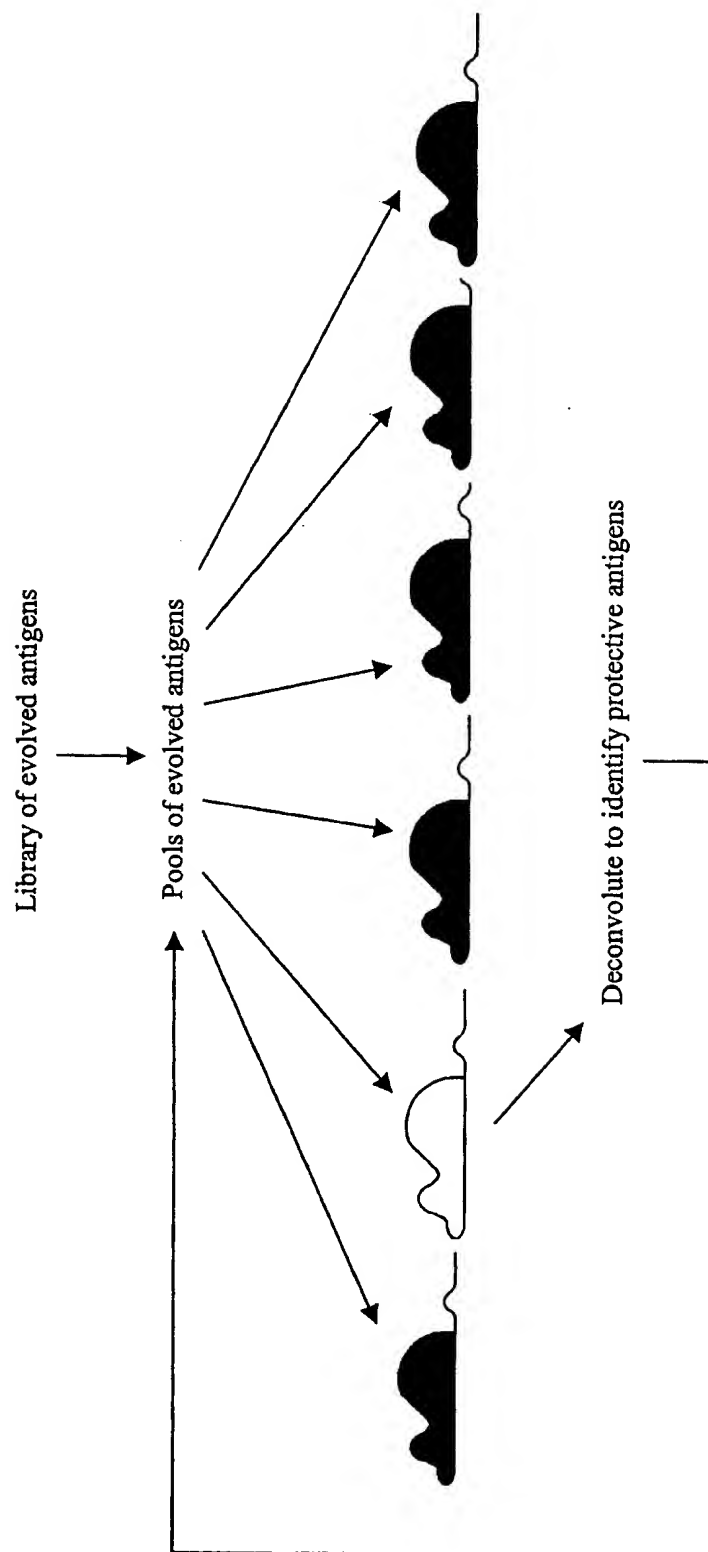
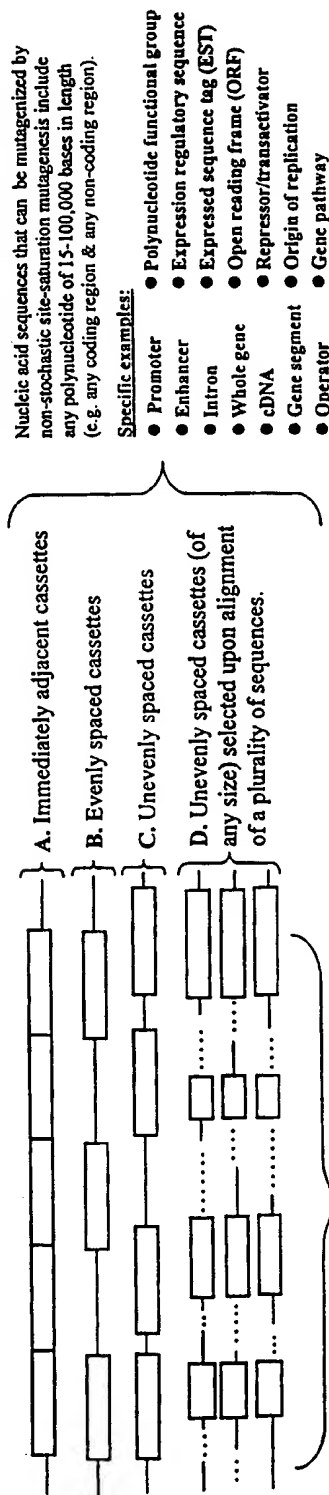


Figure 20. Preferred embodiments of site-saturation mutagenesis

**II. MUTAGENIC CASSETTES WITHIN SEQUENCE TO BE MUTAGENIZED ()**

Mutagenic cassettes that can be mutagenized by non-stochastic site-saturation mutagenesis include any polynucleotide cassette of 1-500 bases in length. Site-saturation mutagenesis is serviceable for mutagenizing a complete set of cassettes contained within a polynucleotide sequence to be mutagenized. As shown, cassettes can be spaced along each polynucleotide differently (i.e. immediately adjacent, evenly spaced, or unevenly spaced) and of any size. In a preferred but non-limiting exemplification a set of mutagenic cassettes is a set of contiguous codons within a sequence of defined length. Alternatively, in another preferred but non-limiting example, a set of mutagenic cassettes is a set of nucleotide cassettes that are not shared by aligned related polynucleotides.

III. TYPES OF MUTATIONS THAT CAN BE INTRODUCED INTO MUTAGENIC CASSETTES

The type of mutations to be introduced in a set of mutagenic cassettes can be of the same type or of different types within each round of polynucleotide site-saturation mutagenesis. Each mutagenic cassette (within the nucleic acid sequence to be mutagenized) preferably is usually mutagenized by the use of a corresponding oligo (including by a degenerate oligo). Examples of degenerate mutations provided by this invention include:

- Codons for all 20 amino acids (e.g. N,N,N or N,N,G/T or N,N,G/C)
- All degenerate codons that do not change the amino acid sequence of the parental template (i.e. codons for the same amino acid that is present in the parental template)
- Codons (all or selected) for amino acids within the same grouping according to the selected amino acid grouping scheme*.
- Codons for at least 1 amino acid in each amino acids group*.

*Exemplary amino acid grouping schemes (notes, some groups overlap each other):

- Aromatic (Phe, Trp, Tyr)
- Aliphatic (Gly, Ala, Val, Leu, Ile)
- OH-containing (Ser, Tyr, Thr)
- Acidic (Asp, Glu, Asn, Gln)
- Basic (Lys, Arg, His)
- Sulfur-containing (Met, Cys)
- Polar (Ser, Thr, Cys, Asn, Gln, Tyr)
- Non-polar (Gly, Ala, Val, Leu, Ile, Met, Phe, Trp, Pro)

Figure 21

**Schematic representation of a multimodule genetic vaccine vector
(relative sizes of functional units are not drawn to scale)**

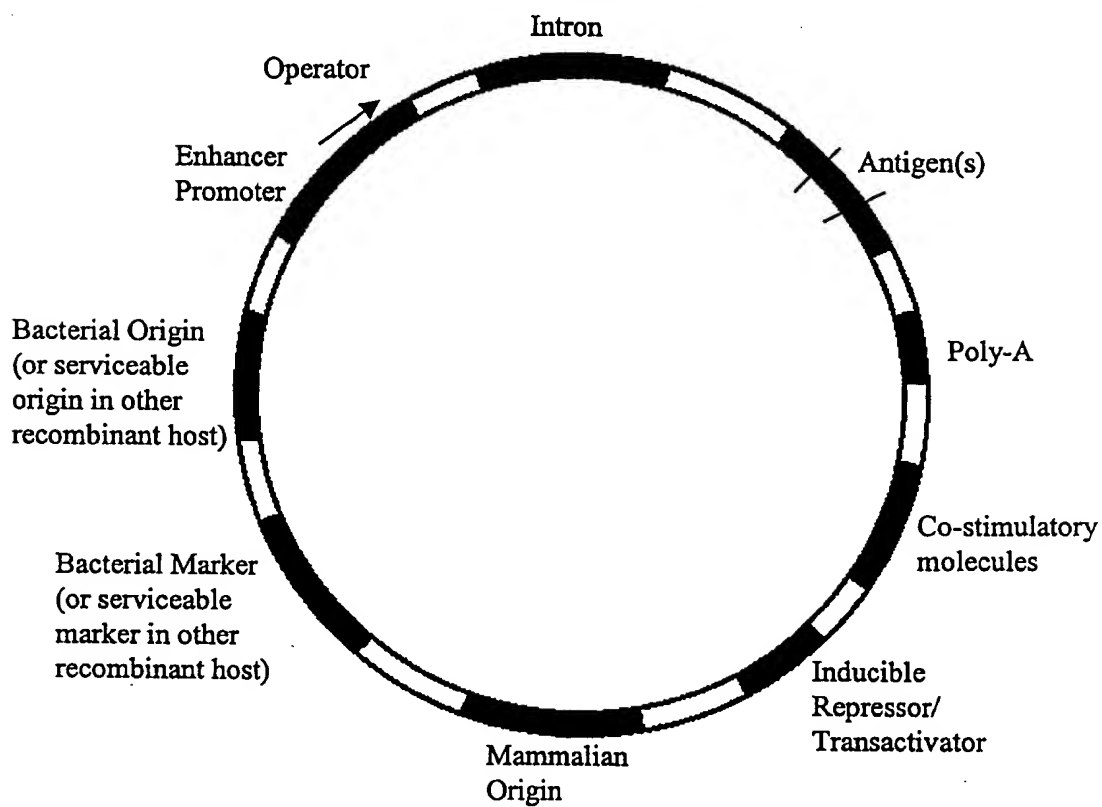


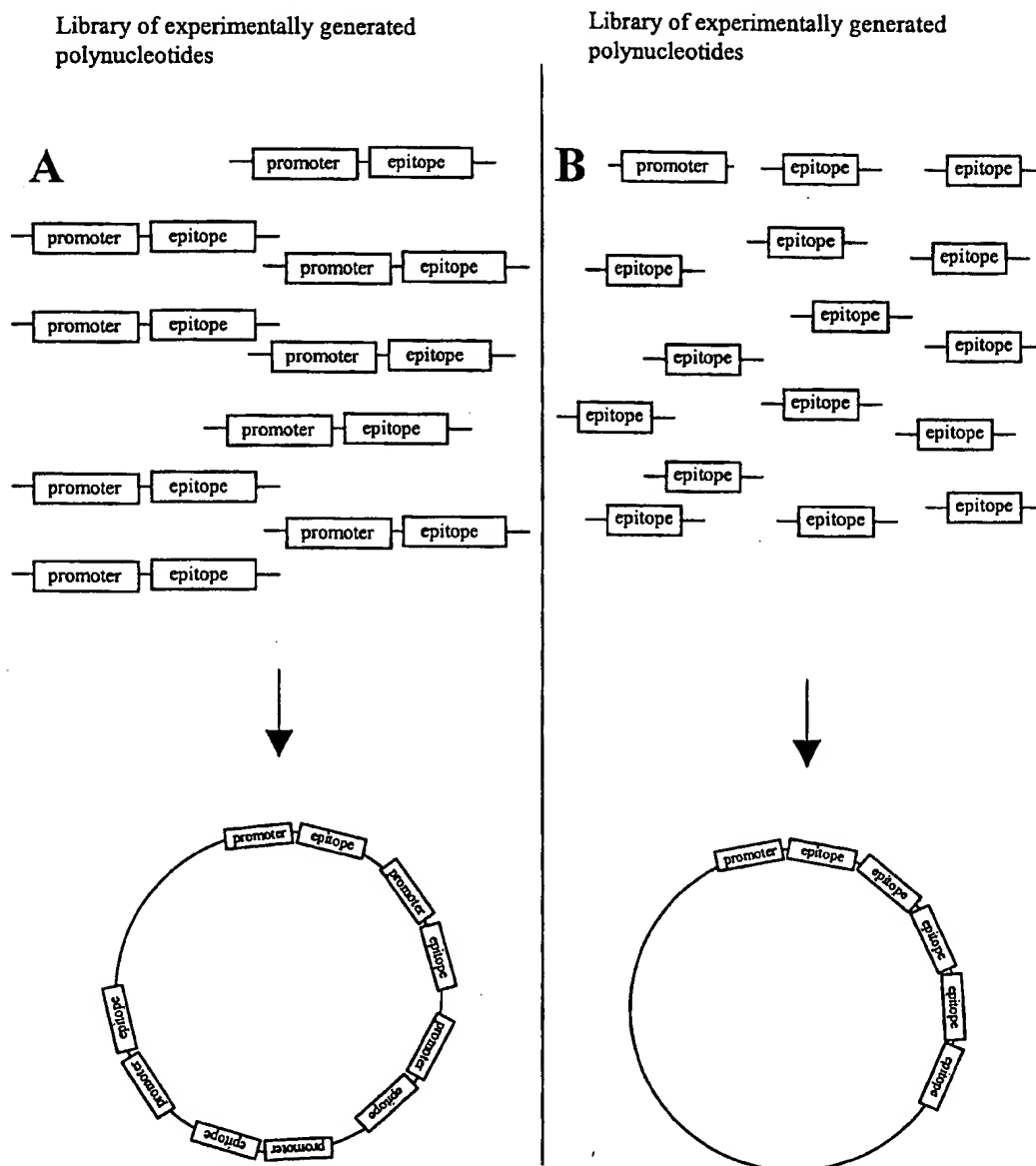
Figure 22A and 22B**Generation of vectors with multiple T cell epitopes.**

Figure 23

Generation of optimized genetic vaccines by directed evolution

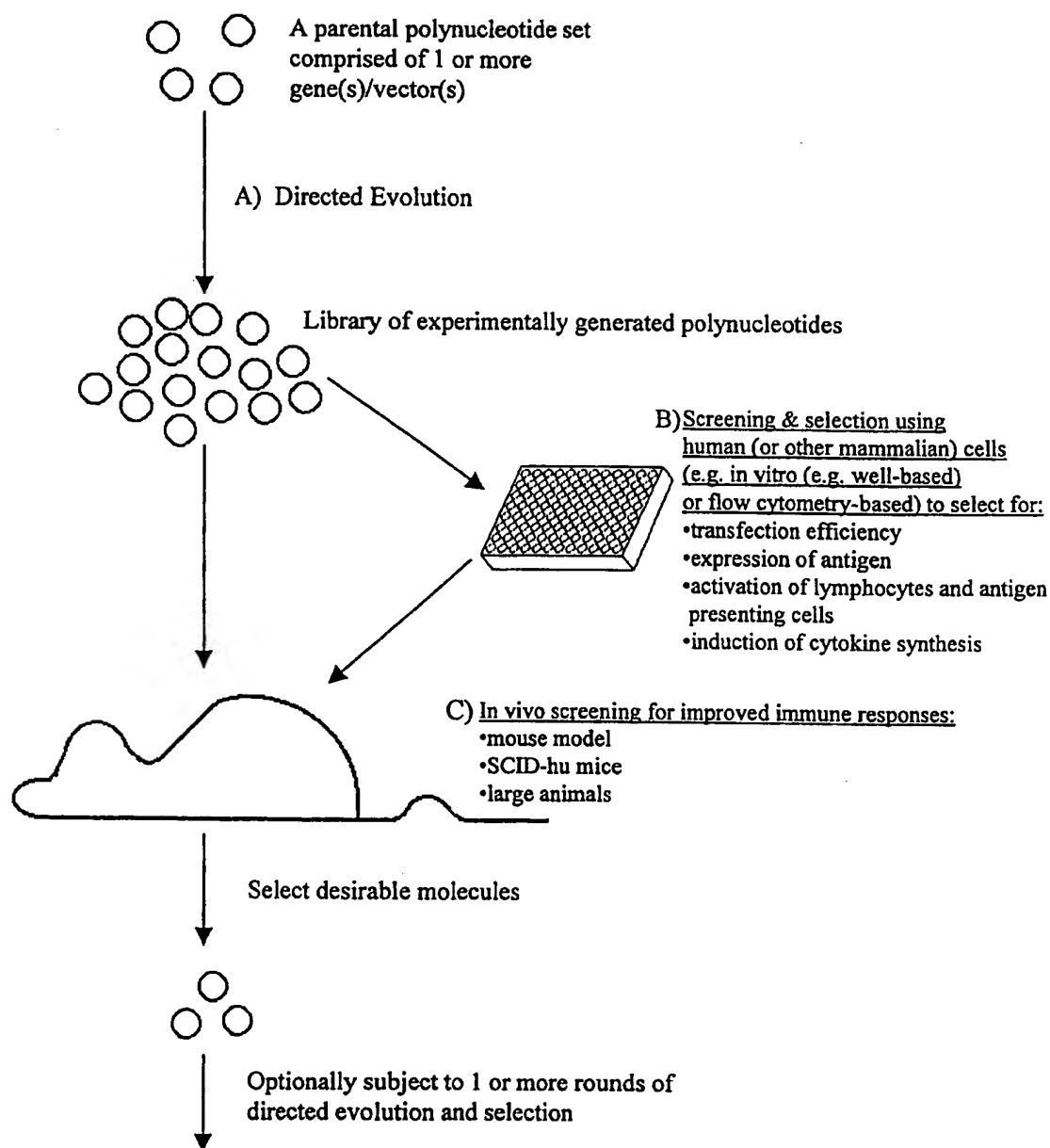


Figure 24

Recursive application of directed evolution and selection of evolved promoter sequences as an example of flow cytometry-based screening methods.

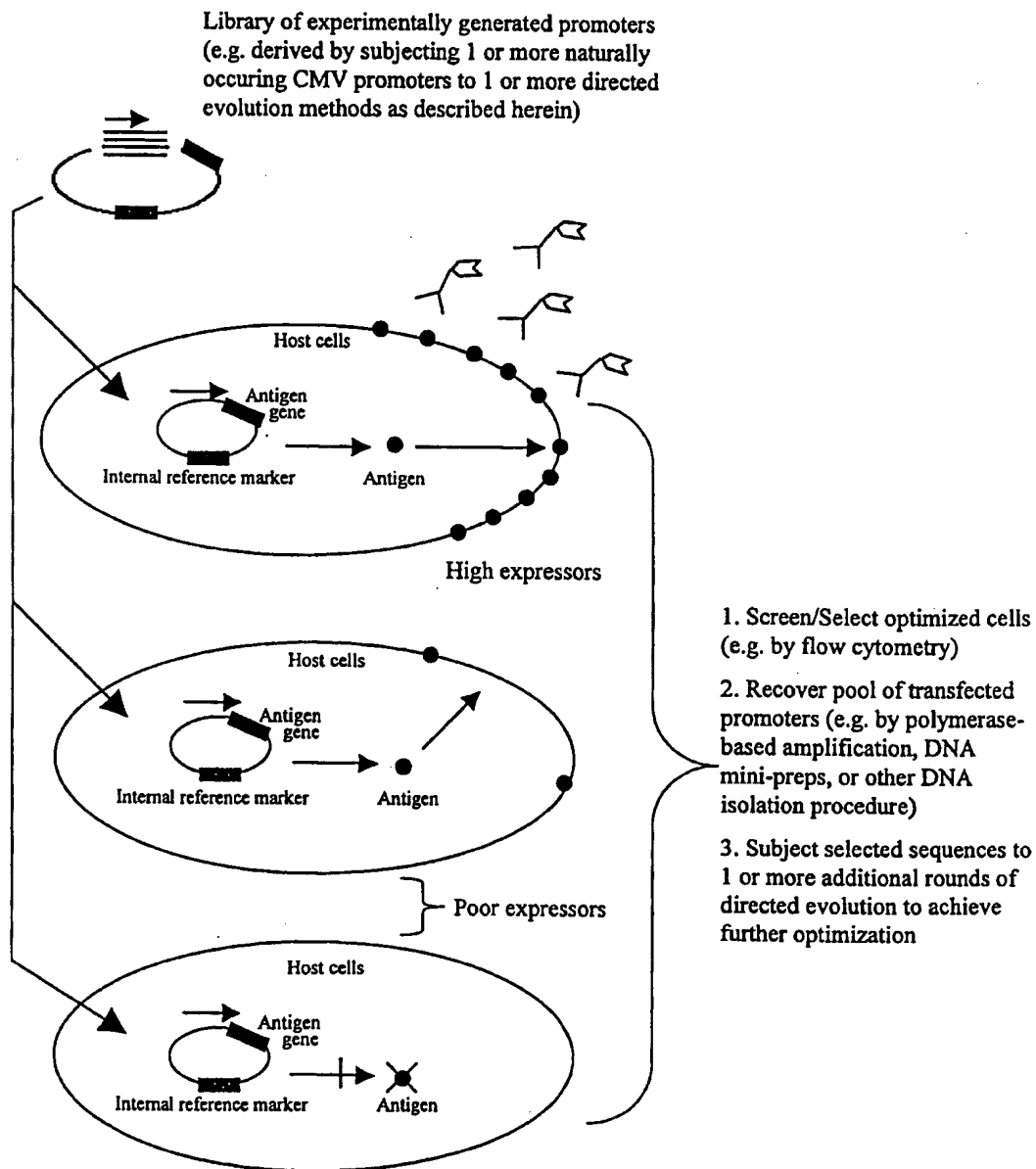


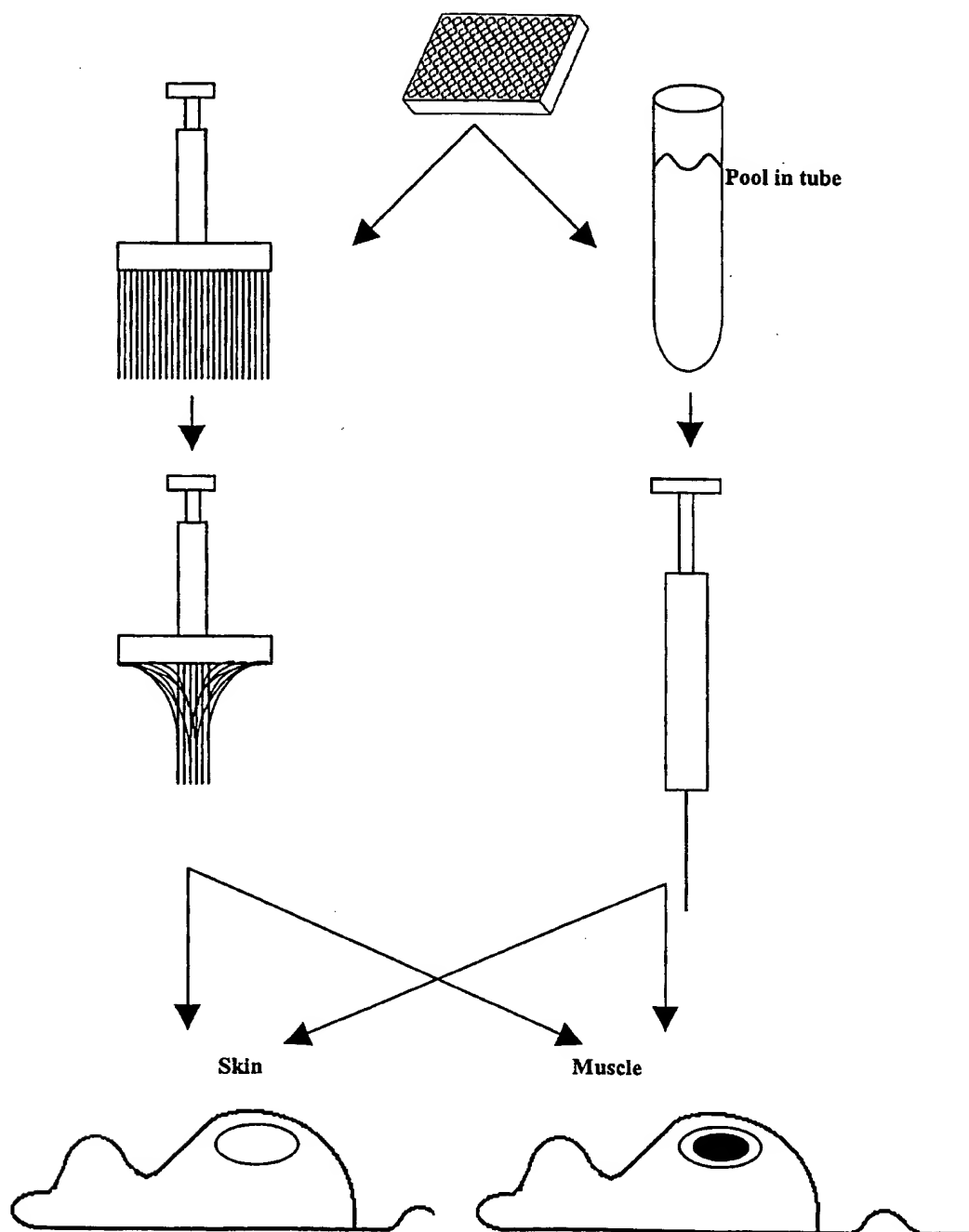
Figure 25**An apparatus for microinjections of skin and muscle.**

Figure 26 Panel A

Non-stochastic polynucleotide reassembly in combination with non-stochastic polynucleotide site-saturation mutagenesis.

Shown below is a non-limiting example of a permutation of the directed evolution methods described herein

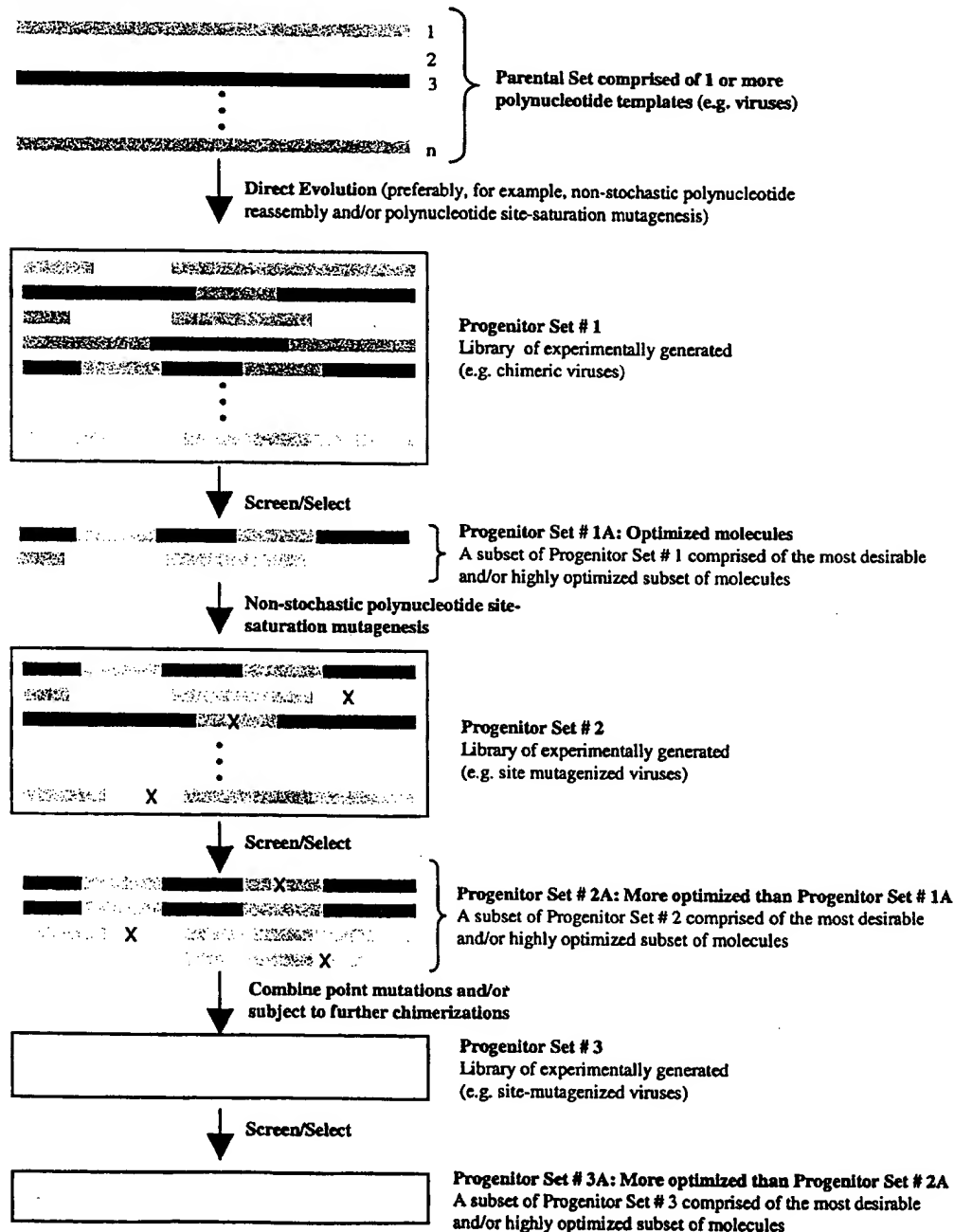


Figure 26 (continued) Panel B

Screening of experimentally generated molecules produced by non-stochastic polynucleotide reassembly in combination with non-stochastic polynucleotide site-saturation mutagenesis

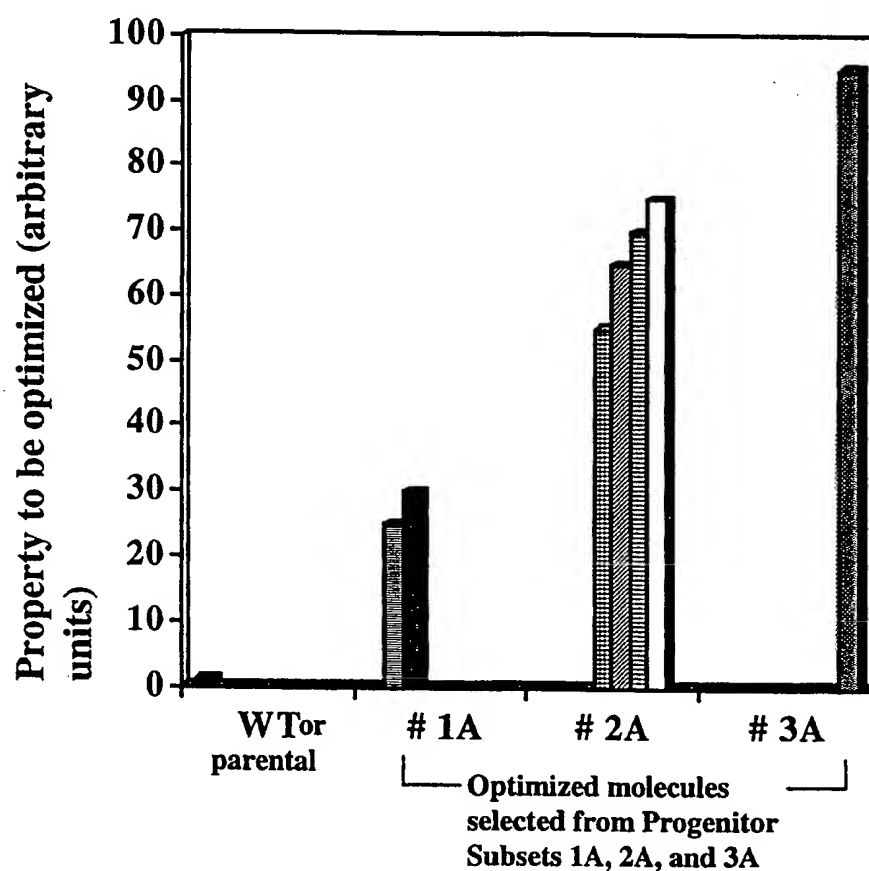


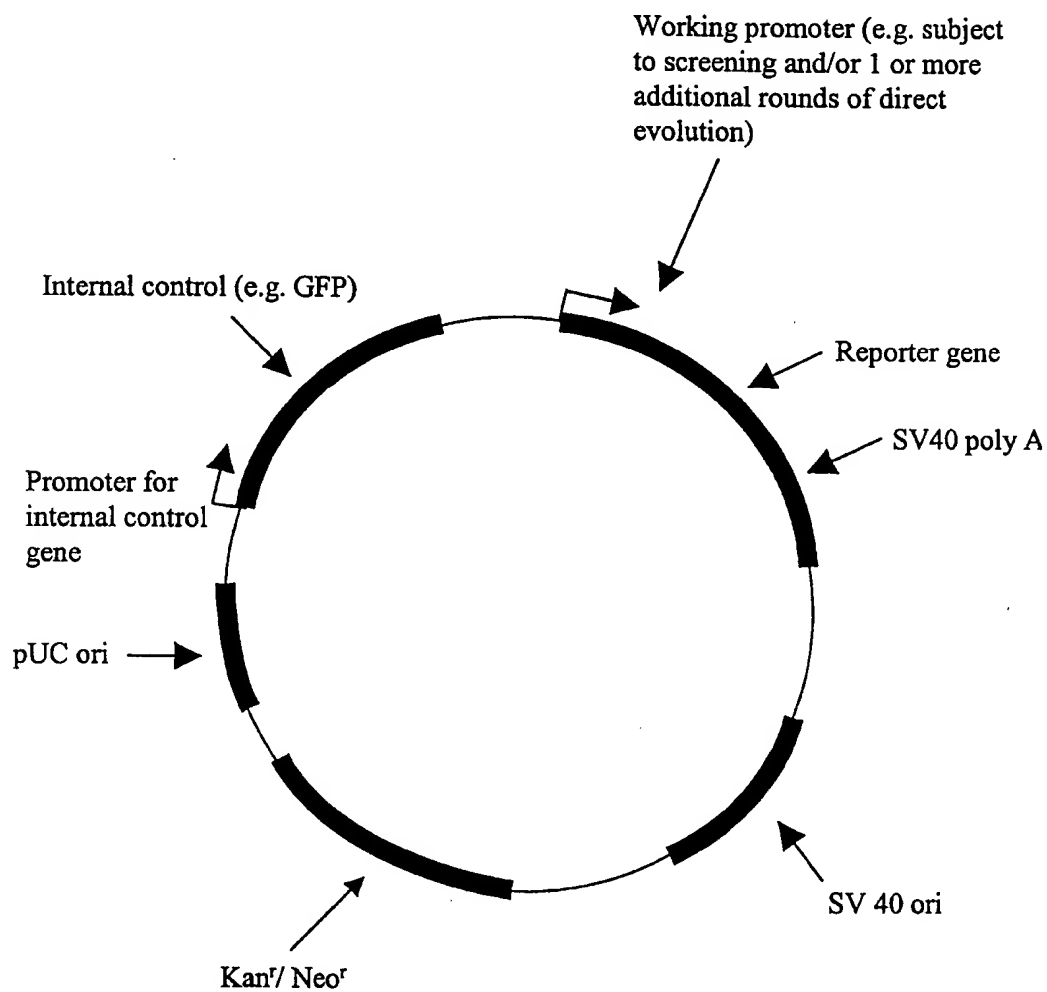
Figure 27**Vector for promoter evolution**

Figure 28

Iterative evolution of inducible promoters using directed evolution and flow cytometry-based selection.

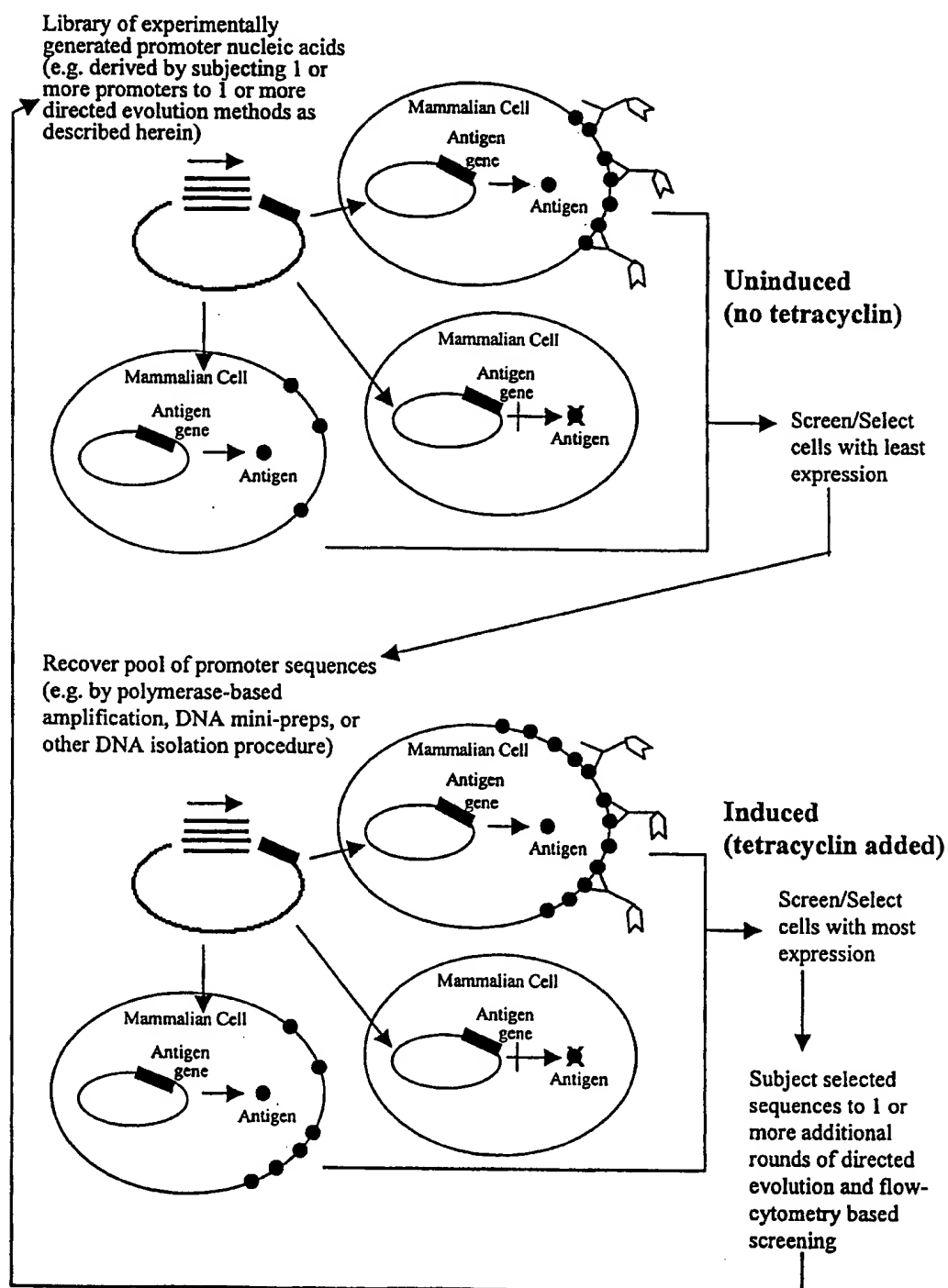
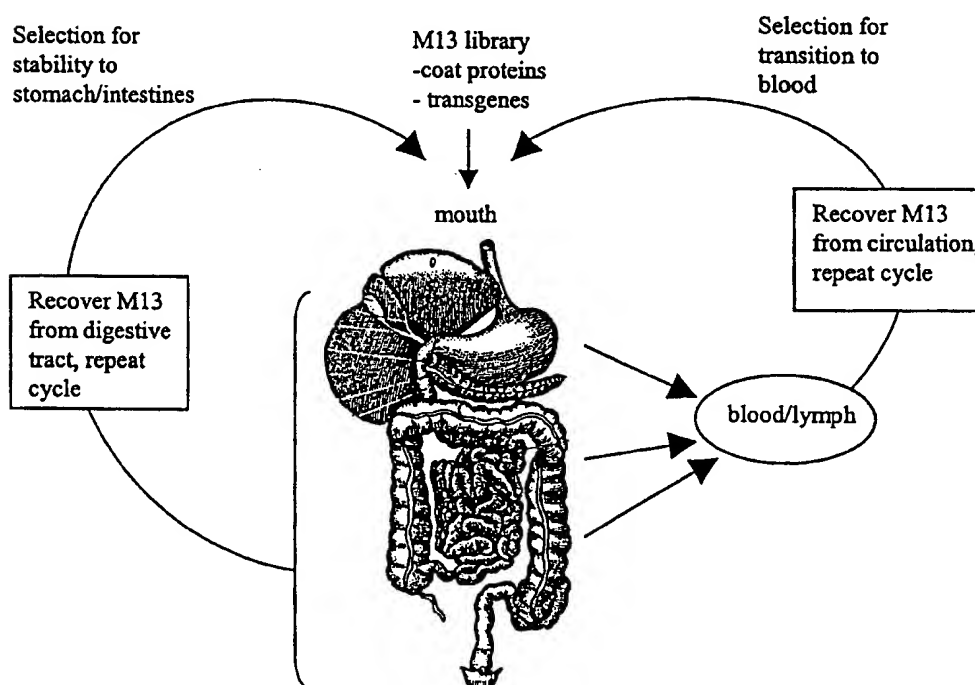


Figure 29

The present invention provides that a genetic vaccine can be subjected to directed evolution in order to achieve improved effectiveness upon administration by oral, intravenous, intramuscular, intradermal, anal, vaginal, or topical delivery methods.

The figure below shows an example of the directed evolution of a genetic vaccine, comprised of an M13 phage-based vaccine, to achieve optimization for oral delivery.



An alignment of the nucleotide sequences of two human CMV strains and one monkey strain.

32 / 50

Figure 30 continued

		551		600
AF026939 CMV	(501)	AAAGGCTGAGAGGGAATTAAGGTCTAAAT		
AF047524 hum UL104	(527)	TTGCAACGCTTCCTCCATGAGCA		
AF078102 Rhesus	(534)	GAACCAAGAAACACCGAGTTGATGATG		
		601		650
AF026939 CMV	(551)	GAAAGCGAGTTTAAAGGTTGGAGAAATGCT		
AF047524 hum UL104	(571)	GGGCGCGCTCTGAGGCTGAGGAGGAGG		
AF078102 Rhesus	(579)	TTCTATGACATGAAAGGCTCTGCTGCTGCT		
		651		700
AF026939 CMV	(597)	ACAACCTCAATCTCTCCCTACGGAATGATGACCTCA		
AF047524 hum UL104	(615)	GGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(627)	TTTAAAGCAAAATGTTTCTGCGAGGAGCTGCTGCT		
		701		750
AF026939 CMV	(646)	TATACCTAAACCTTCTCTCTTCTTAAAGG		
AF047524 hum UL104	(657)	CTCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(677)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
		751		800
AF026939 CMV	(689)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF047524 hum UL104	(706)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(725)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
		801		850
AF026939 CMV	(736)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF047524 hum UL104	(756)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(775)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
		851		900
AF026939 CMV	(784)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF047524 hum UL104	(801)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(821)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
		901		950
AF026939 CMV	(834)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF047524 hum UL104	(847)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(863)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
		951		1000
AF026939 CMV	(884)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF047524 hum UL104	(896)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(912)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
		1001		1050
AF026939 CMV	(934)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF047524 hum UL104	(942)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(961)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
		1051		1100
AF026939 CMV	(984)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF047524 hum UL104	(986)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(1004)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
		1101		1150
AF026939 CMV	(1032)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF047524 hum UL104	(1036)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		
AF078102 Rhesus	(1049)	CTCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT		

AF026939 CMV	(1082)	1151	1200
AF047524 hum UL104	(1084)		
AF078102 Rhesus	(1095)		
AF026939 CMV	(1131)	1201	1250
AF047524 hum UL104	(1131)		
AF078102 Rhesus	(1143)		
AF026939 CMV	(1175)	1251	1300
AF047524 hum UL104	(1178)		
AF078102 Rhesus	(1175)		
AF026939 CMV	(1222)	1301	1350
AF047524 hum UL104	(1226)		
AF078102 Rhesus	(1222)		
AF026939 CMV	(1266)	1351	1400
AF047524 hum UL104	(1274)		
AF078102 Rhesus	(1267)		
AF026939 CMV	(1315)	1401	1450
AF047524 hum UL104	(1323)		
AF078102 Rhesus	(1317)		
AF026939 CMV	(1361)	1451	1500
AF047524 hum UL104	(1367)		
AF078102 Rhesus	(1356)		
AF026939 CMV	(1411)	1501	1550
AF047524 hum UL104	(1413)		
AF078102 Rhesus	(1403)		
AF026939 CMV	(1457)	1551	1600
AF047524 hum UL104	(1462)		
AF078102 Rhesus	(1451)		
AF026939 CMV	(1506)	1601	1650
AF047524 hum UL104	(1508)		
AF078102 Rhesus	(1501)		
AF026939 CMV	(1552)	1651	1700
AF047524 hum UL104	(1553)		
AF078102 Rhesus	(1547)		
AF026939 CMV	(1600)	1701	1750
AF047524 hum UL104	(1602)		
AF078102 Rhesus	(1587)		

Figure 30 continued

		1751	1800
AF026939 CMV	(1649)	CT---TG---TCTAAGGAGGTTTATTCCTCTAGGAGG	
AF047524 hum UL104	(1644)	GG---CCTCTC---G---TCCAC---CAGAT	
AF078102 Rhesus	(1632)	TTCAAGAGTAA---AAAGAGAA---CG---T---GAGG	
		1801	1850
AF026939 CMV	(1696)	TTTTTAAAGTGTTTTGTGACCGGCTTGTCTGAGCTT---	
AF047524 hum UL104	(1686)	CCCCGGTCC---CAAGGCGT---GAGAGGCGCGCGC	
AF078102 Rhesus	(1676)	AA---TTTTTAAAGGCTGAAG---CTTAT---AA	
		1851	1900
AF026939 CMV	(1744)	TC---CA---TGGAGGC---AGGAGGATCACAGGCTGGAAT-	
AF047524 hum UL104	(1736)	AGCTCAGCTCCGCCCA---CCTCCTCTAAGTTCAAGAGG	
AF078102 Rhesus	(1718)	AGGAGGTTGAGT---GAGAGGAGGTTCAATTTTCC	
		1901	1950
AF026939 CMV	(1793)	TGAGCTCTCTTGGTACGAGGAGATCTCTCTCTAAATATA	
AF047524 hum UL104	(1786)	TTTCAAGCTCTAGGAGGCTTGGCGTCTCTCTCTCTCTCT	
AF078102 Rhesus	(1760)	TT---CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	
		1951	2000
AF026939 CMV	(1843)	AA---AAT---GCGAGGAGTGGCTGCTCTCTCTCTCTCTCT	
AF047524 hum UL104	(1836)	GCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	
AF078102 Rhesus	(1809)	TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	
		2001	2050
AF026939 CMV	(1893)	AGCTGTGGGAGGTA---GATGGCGTGAAGCTGAAGGAGGCTT	
AF047524 hum UL104	(1882)	CAGGACCGGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCT	
AF078102 Rhesus	(1853)	---GAGGAGAGTT---GATATTAGAGACCT-ATCTCTCT---	
		2051	2100
AF026939 CMV	(1940)	TCGCAAGATT---GGGCTCTG---ATCTGCTCTCTCTCTCT	
AF047524 hum UL104	(1931)	ACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	
AF078102 Rhesus	(1892)	AAAGCTGAGTCT---AGATG---ATGAAATTCTCTCTCTCT	
		2101	2150
AF026939 CMV	(1984)	AGGACTTGAATTCCTGAGCGGGGCTCTCT---TCTCTTGGCC	
AF047524 hum UL104	(1981)	TTTCTTGAACCTTGGTGGGCTCTCTCTCTCTCTCTCTCT	
AF078102 Rhesus	(1932)	GGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	
		2151	2200
AF026939 CMV	(2032)	GGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	
AF047524 hum UL104	(2031)	TTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	
AF078102 Rhesus	(1979)	TTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	
		2201	2214
AF026939 CMV	(2057)	-----	
AF047524 hum UL104	(2081)	TGGTTTCGCTCCAT	
AF078102 Rhesus	(1999)	-----	

Figure 31

An alignment of IL-4 nucleotide sequences from 3 species
(human, primate, and canine).

AF187322 Canis IL-4	(1)	1	50
NM_000589 Homo sapien IL-4	(1)		
U19838 Cercopithecus IL-4	(1)		
AF187322 Canis IL-4	(51)	51	100
NM_000589 Homo sapien IL-4	(50)		
U19838 Cercopithecus IL-4	(1)		
AF187322 Canis IL-4	(101)	101	150
NM_000589 Homo sapien IL-4	(100)		
U19838 Cercopithecus IL-4	(35)		
AF187322 Canis IL-4	(151)	151	200
NM_000589 Homo sapien IL-4	(150)		
U19838 Cercopithecus IL-4	(85)		
AF187322 Canis IL-4	(201)	201	250
NM_000589 Homo sapien IL-4	(199)		
U19838 Cercopithecus IL-4	(134)		
AF187322 Canis IL-4	(250)	251	300
NM_000589 Homo sapien IL-4	(249)		
U19838 Cercopithecus IL-4	(184)		
AF187322 Canis IL-4	(300)	301	350
NM_000589 Homo sapien IL-4	(299)		
U19838 Cercopithecus IL-4	(234)		
AF187322 Canis IL-4	(319)	351	400
NM_000589 Homo sapien IL-4	(349)		
U19838 Cercopithecus IL-4	(284)		
AF187322 Canis IL-4	(346)	401	450
NM_000589 Homo sapien IL-4	(399)		
U19838 Cercopithecus IL-4	(334)		
AF187322 Canis IL-4	(393)	451	500
NM_000589 Homo sapien IL-4	(449)		
U19838 Cercopithecus IL-4	(384)		
AF187322 Canis IL-4	(443)	501	550
NM_000589 Homo sapien IL-4	(499)		
U19838 Cercopithecus IL-4	(434)		

Figure 31 continued

[illegible]

Figure 32

Evolution of polypeptides by synthesizing (in vivo or in vitro) corresponding deduced polynucleotides and subjecting the deduced polynucleotides to directed evolution and expression screening subsequently expressed polypeptides.

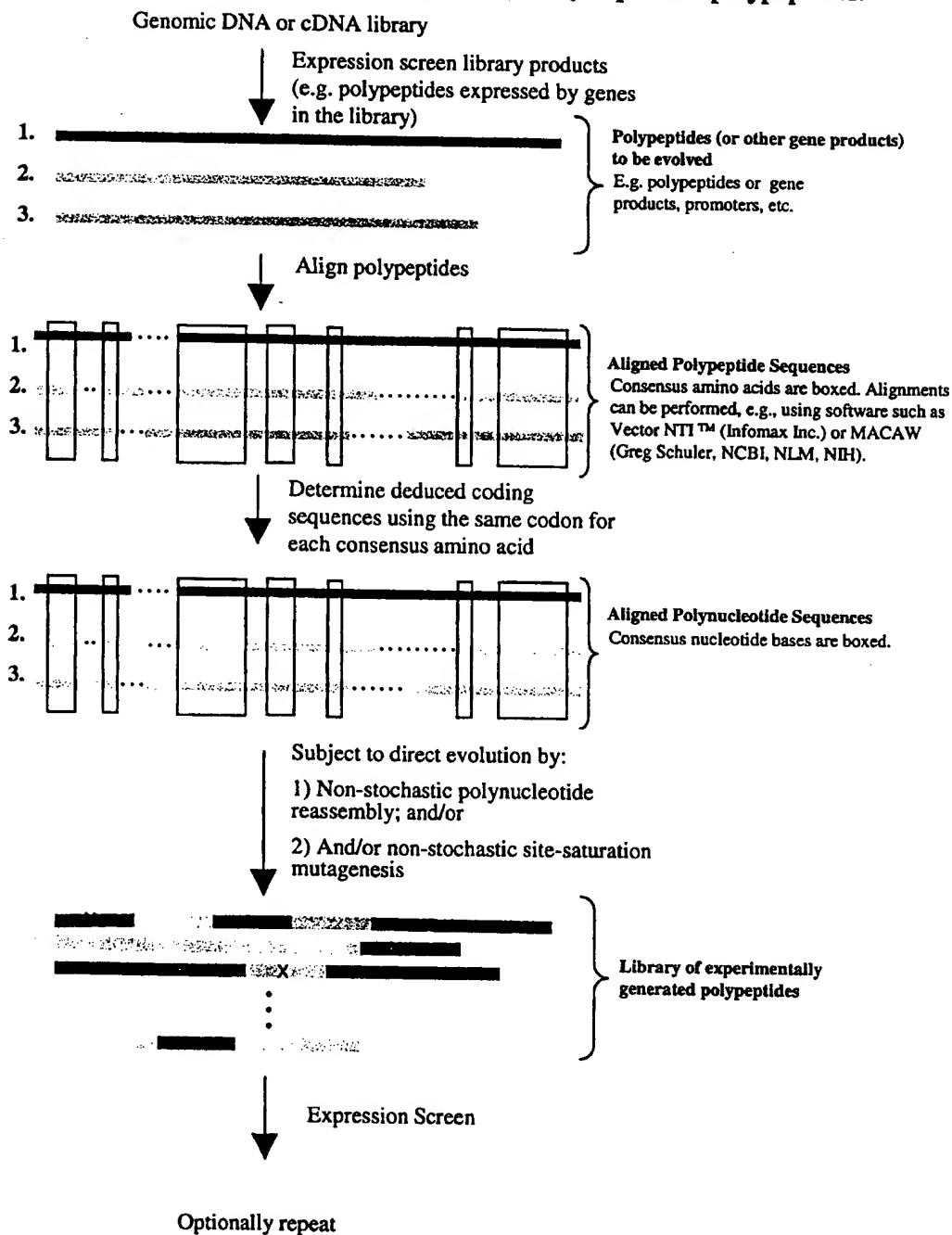


Figure 33

Directed evolution of polynucleotides (e.g. promoter sequences)

This figure shows an example of the application of non-stochastic site-saturation mutagenesis in combination with non-stochastic reassembly (e.g. oligo-directed CpG deletion(s) and/or addition(s))

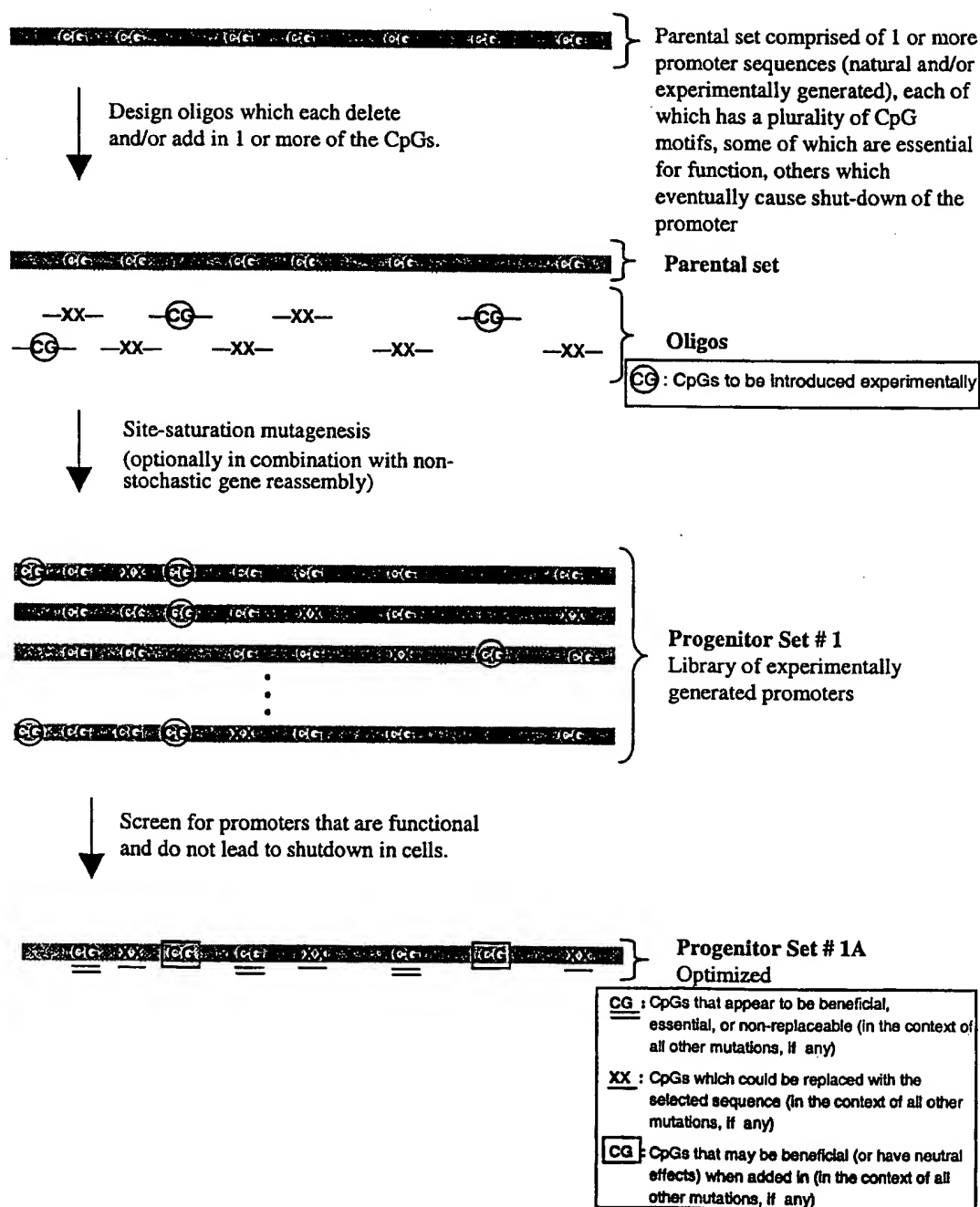


Figure 34

An example of a CTIS obtained from HbsAg polypeptide (PreS2 plus S regions).

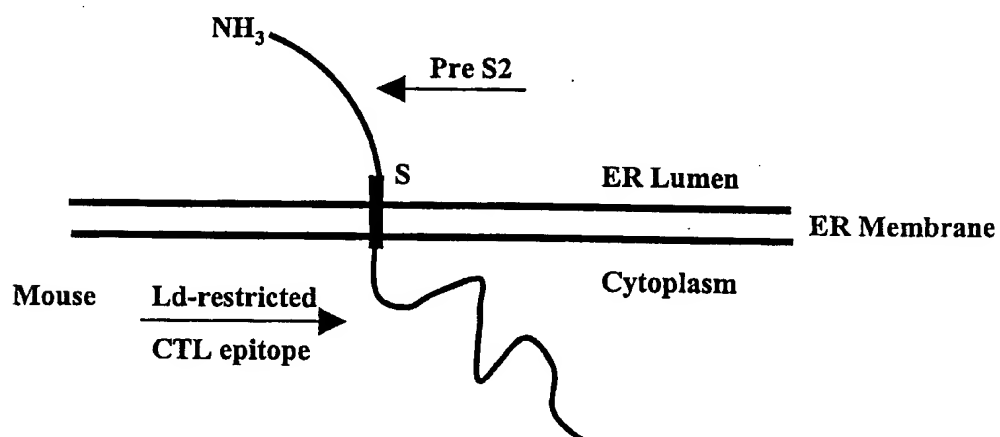


Figure 35

An example of a CTIS having heterologous epitopes attached to the cytoplasmic portion.

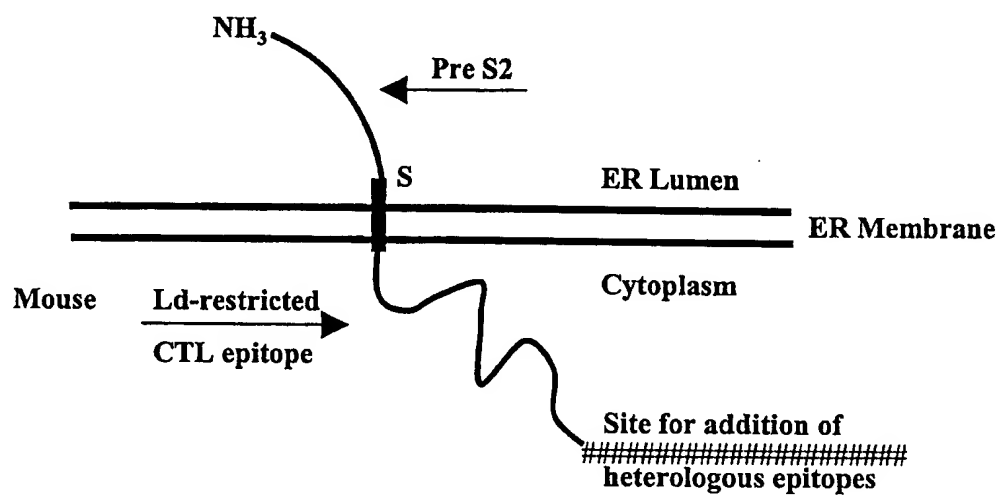


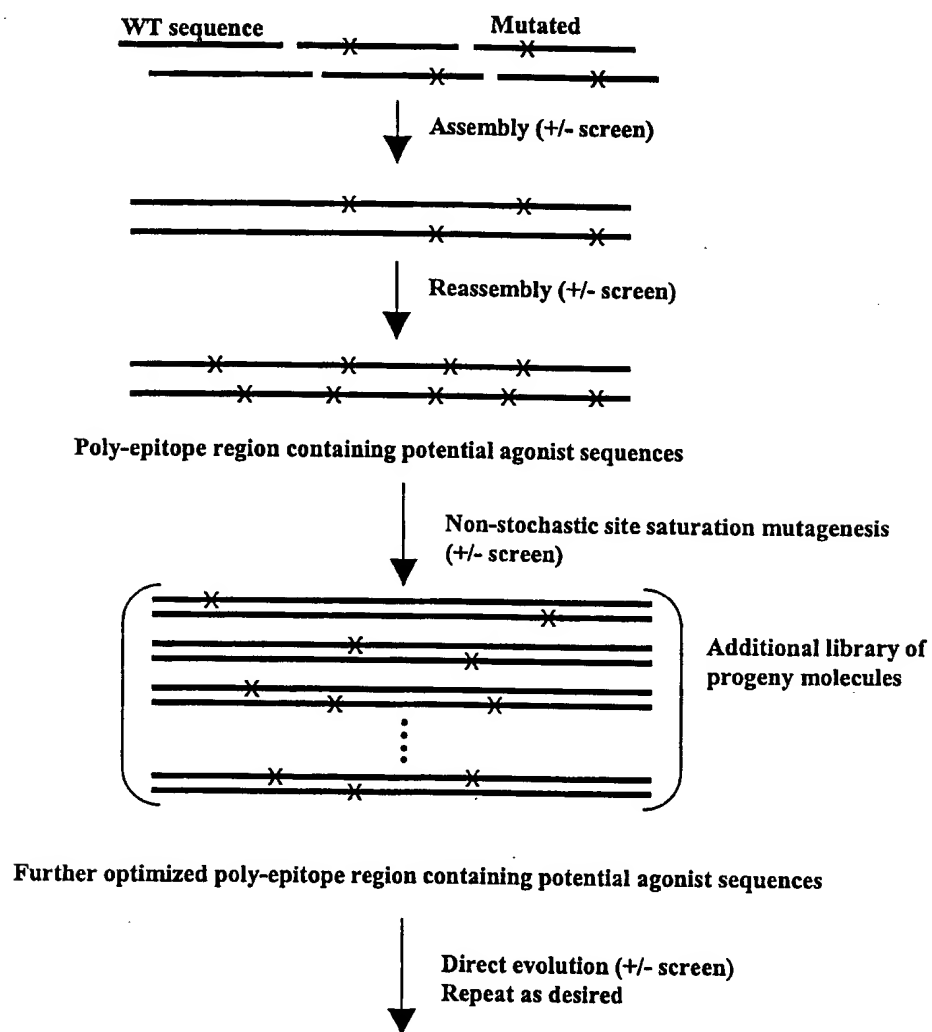
Figure 36**Method for preparing immunogenic agonist sequences (IAS).**

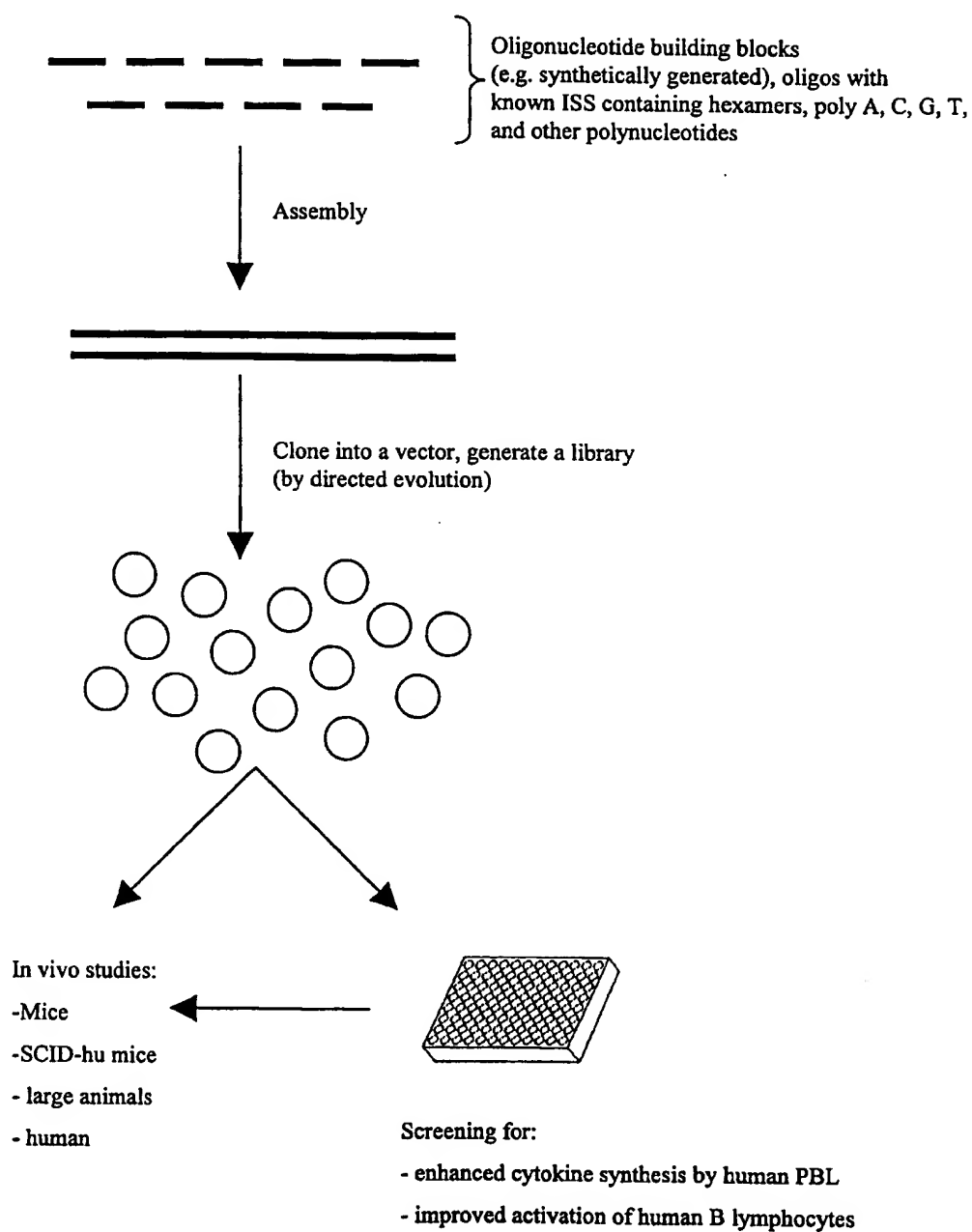
Figure 37**Improving Immunostimulatory Sequences (ISS) Using Directed Evolution.**

Figure 38

Screening to identify IL-12 genes that encode recombinant IL-12 having an increased ability to induce T Cell proliferation.

Working Progenitor templates

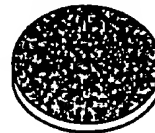
Library of IL-12 genes
(p35/p40 fusions)

1) Directed Evolution



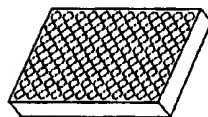
2) Express in bacterial host

Bacterial colonies

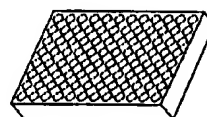


8) Optionally repeat steps 1-7

7) Identification and selection
of clones inducing most potent
T cell proliferation

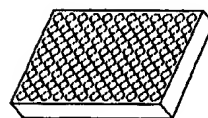


3) Robotic colony picking
(one colony/well)

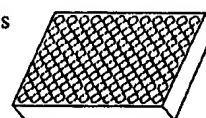


96 wells X 50

6) Transfer of supernatants
to human T cell cultures



4) High throughput plasmid purification,
(e.g. PERFECT prep-96 kit)



5) Transfection to CHO cells



Figure 39

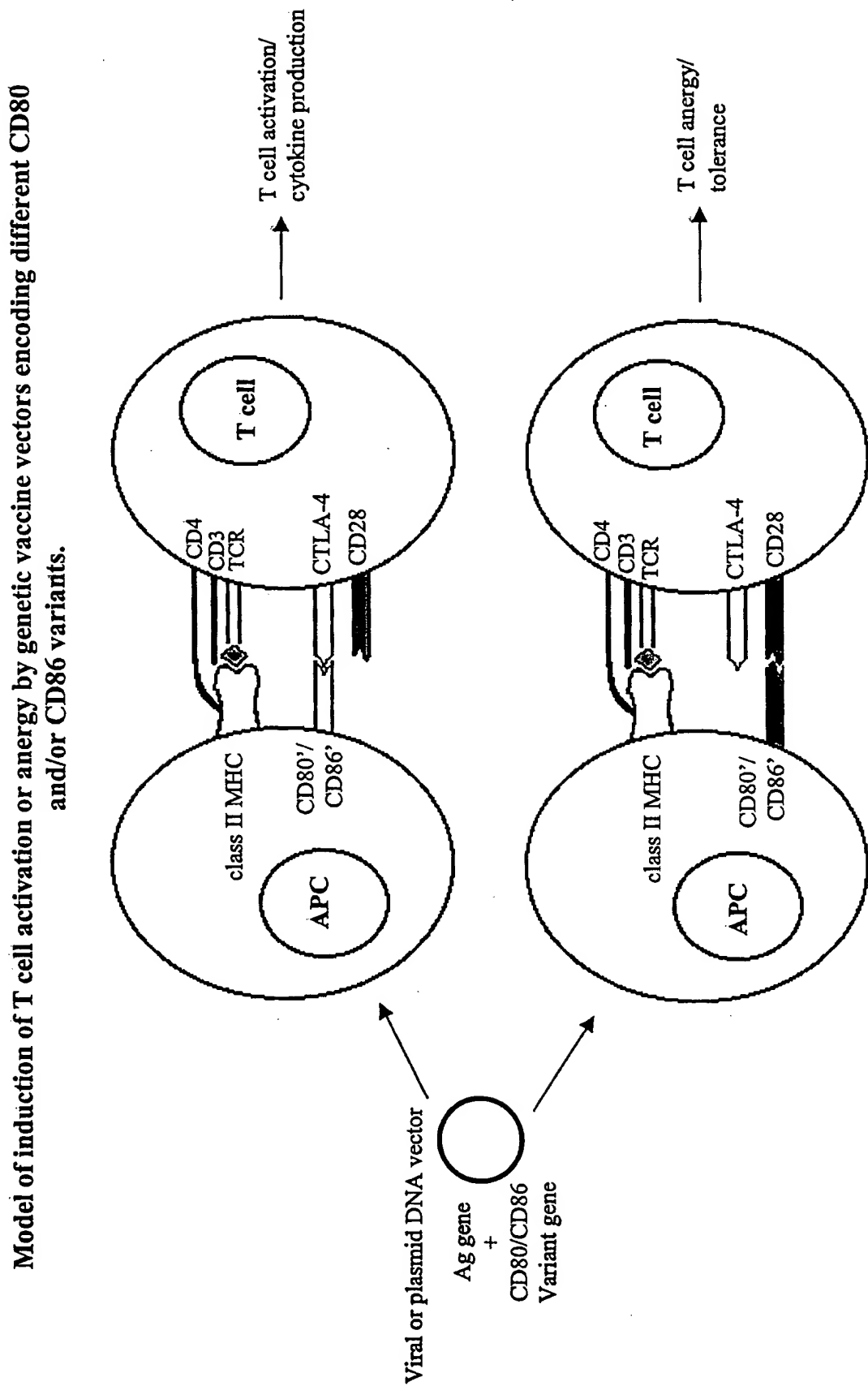


Figure 40

Screening to identify CD80/CD86 chimeric genes having an improved capacity to induce T Cell activation or anergy.

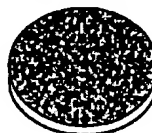
1) Directed Evolution

Library of Working
Progenitor templates of
CD80 &/or CD86 genes

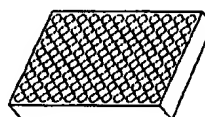


2) Express in bacterial host

Bacterial colonies

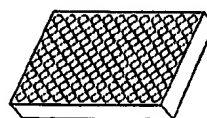


3) Robotic colony picking
(one colony/well)

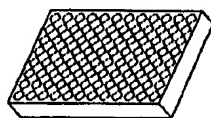


96 wells X 50

4) High throughput vector purification,
(e.g. PERFECT prep-96 kit)



5) Transfection to dendritic cells/
U937 cells



6) Co-culture with
T cell cultures

7) Identification and selection
of clones inducing most potent
T cell activation or anergy

8) Optionally repeat steps 1-7

Figure 41. An alignment of two CMV-derived nucleotide sequences from human and primate species.

47 / 50

Figure 41 continued

AF078102 Rhesus M67443 Towne	(682) (620)	701 [AA][ACC][CAGA][AGTTT][TGG][G][T][C][AGTAG][A][AA][C][C]	750 [--][CAG][TGAT][GTGAC][GTA][G][C][G][GTACC][G][GT][TT]
		751 [GG][CAT][GAG][GATCC][TT][C][C][C][G][--][ACT][TGA][AAG]	800 [CA][GAG][T][C][TCCGG][AA][G][T][A][C][C][C][CT][G][CTC]
AF078102 Rhesus M67443 Towne	(779) (717)	801 [CT][CA][TGTAT][CTTT][CA][AAT][AG][G][T][AT][A][T][AG][AA]	850 [TG][AG][GG][GAGGA][TGAC][A][G][CCC][CA][C][C][C][C][CA][GC]
		851 [T][T][--][TA][AGTAT][TTAA][CA][C][CATC][A][TG][G][--][G][G]	900 [G][C][CC][CG][GCGCA][CGGC][TT][G][GTG][T][G][T][C][AAAATA][G][G]
AF078102 Rhesus M67443 Towne	(871) (817)	901 [--][GT][TT][TGTT][G][G][G][G][G][T][C][AAT][T][G][AA][T][TG][G]	950 [AT][AA][CG][CAAG][T][C][A][A][C][T][G][G][G][G][C][T][T][AC][C]
		951 [CT][T][G][A][AA][CT][T][GAC][A][G][GGC][T][TG][TAT][T][G][G]	1000 [GAC][A][C][T][--][GG][C][CT][T][C][CA][A][CA][CCC][G][C][C]
AF078102 Rhesus M67443 Towne	(966) (914)	1001 [GAT][GAC][AA][G][GACG][T][T][--][GCT][AGC][G][AAG][A][GAA][T][T]	1050 [C][AT][T][CAGT][A][CTAT][GAT][GA][G][GGC][A][GAT][--][G][T][G][T][GAG]
		1051 [G][TTCAG][T][G][GCTT][T][TAA][CT][G][G][C][T][A][G][CG][T][TG][C][GC]	1100 [G][C][--][A][GGA][--][AGCGGA][A][G][G][A][C][G][C][T][G][T][--][G][--]
AF078102 Rhesus M67443 Towne	(1064) (1000)	1101 [G][AT][G][G][G][GAG][G][G][G][G][--][G][G][G][T][G][T][G][A][T][G][T]	1150 [G][AT][C][G][G][G][T][G][G][G][T][T][T][T][T][T][T][T][C][T][C][C][T][G][G][G][G]
		1151 [G][AG][T][G][C][GTA][GCA][G][T][T][T][T][G][G][G][T][T][G][T][T][AC][CA][T]	1200 [G][G][G][G][G][C][T][AGT][CAGC][A][A][A][C][CACC][T][A][C][--][C][G][T][T][C]
AF078102 Rhesus M67443 Towne	(1159) (1096)	1201 [TA][G][G][C][A][G][GCT][C][T][G][T][G][G][T][--][G][A][G][G][GCT][G][G][G][T][AT]	1250 [AT][G][A][--][G][G][G][AGCT][GAGTAC][GAC][GAC][G][G][G][--][G][G][G][G][G][G]
		1251 [C][G][C][T][ATGAT][AAATAC][TAC][AGT][G][T][T][T][G][G][GACG][TAT][GAT][T]	1300 [AG][G][T][G]
AF078102 Rhesus M67443 Towne	(1258) (1192)	1301 [G][T][CCT][C][GGA][G][T][T][GCA][A][G][AATC][A][A][T][T][G][G][T][G][T][G][G]	1350 [G][C][AGG][A][TC][--][ACCACC][A][C][GCA][A][G][G][G][G][G][G][G][G][G][G]
		1351 [G][AT][T][G][C][T][TCTA][AAG][G][G][--][G][AAG][G][T][C][T][G][G][G][--][G][A]	1400 [G][G][C][C][T][G][--][G][G][C][C][T][G][T][T][C][C][C][C][C][C][C][C][C][C]
AF078102 Rhesus M67443 Towne	(1308) (1241)	1351 [G][AT][T][G][C][T][TCTA][AAG][G][G][--][G][AAG][G][T][C][T][G][G][G][--][G][A]	1400 [G][G][C][C][T][G][--][G][G][C][C][T][G][T][T][C][C][C][C][C][C][C][C][C][C]
		1351 [G][AT][T][G][C][T][TCTA][AAG][G][G][--][G][AAG][G][T][C][T][G][G][G][--][G][A]	1400 [G][G][C][C][T][G][--][G][G][C][C][T][G][T][T][C][C][C][C][C][C][C][C][C][C]

Figure 41 continued

AF078102 Rhesus	(1355)	1401	1450
		ATGCTGAGTAAAGAAAGAACTGATGTACTTGAATTTCTTT	
M67443 Towne	(1286)	ATCAGC--CCTCCCGGAGGCACGCGGGCTTAAACACGCG	
AF078102 Rhesus	(1405)	1451	1500
		A-AAGAGGTTTTCACAGAGGAAATTTGGGTGAAGGCT	
M67443 Towne	(1335)	CCCTTTAAGCCGAGAG-----AGGTCGCCCGAAGAGTCTCGTC	
AF078102 Rhesus	(1454)	1501	1550
		TACTCAATAGTGGGATGATGTTCCTATGAAGAAGGCC	
M67443 Towne	(1381)	AGGTTTCCTCACT--ATCC-----CTATCGGCTGTTCAGTG	
AF078102 Rhesus	(1504)	1551	1600
		TAAATCACTTTTATGCTTAAATGAACAAACGGAGTATG	
M67443 Towne	(1425)	GCCCTGTGAGCCGATCCGGCGCACTGGTGCCATGTATG	
AF078102 Rhesus	(1554)	1601	1650
		ATTGTTAGTTAAGAGGTGATCA--TCTG--CCG--GAGATACGA	
M67443 Towne	(1475)	CTAAGGTCTCGGCTGAAGCTAAGAACAGAGTCTTTGGGCTC	
AF078102 Rhesus	(1600)	1651	1700
		TGATCCGCAACCAGAGTTATGGGATATCCAAAGCAATGCT--	
M67443 Towne	(1525)	--AAGGATGTACG--GATCTCCGGATGGAGGGCTATGAGGC	
AF078102 Rhesus	(1648)	1701	1750
		-----AAGGAGACCGTTACCAATGAAGGAATCTTATCTC	
M67443 Towne	(1571)	CCGCTGCGCAACCAACCTGCGCCACCGCAAGACGCC--TGCCCG	
AF078102 Rhesus	(1689)	1751	1800
		GGGCGG---GAGGCTATACATGAAAGATGGCATTGTTCT	
M67443 Towne	(1620)	CCAAAGTCTGCTGACCCCAAGACACCGGCTTGAAGCCACCCTC	
AF078102 Rhesus	(1735)	1801	1850
		CGAAAGGGTACG-----TTATGACCTTTAATGATATATTAGT	
M67443 Towne	(1670)	CCGCCTCTAGGCGACTCTATGAAAGGACCTCCATCTAGACAGCG	
AF078102 Rhesus	(1779)	1851	1900
		GGTGGATTAAAGATGCTTTGATTGAATATCTAATCTGGT--CTA	
M67443 Towne	(1720)	AGTTTGGCCGCA---GACCTTCGCGGCTATATTTGGACAGTTG	
AF078102 Rhesus	(1827)	1901	1950
		CCGTGTATTTCTCTAGTGGTGGCAATGAATATATTTAGAT	
M67443 Towne	(1767)	CCGGAACCTTCCG---CTCCACCAAGACCGAT--CACCTTCCGA	
AF078102 Rhesus	(1877)	1951	2000
		-GAGTATCAATTGTGAAAGCTGCTCAATCATGCATTTGGGAGGA	
M67443 Towne	(1814)	TCCCTGACCCCC-----CTCTCCCGCTCGCGATGCTCAGGAT	
AF078102 Rhesus	(1926)	2001	2050
		GATCGGGCTTTCTTTCTTTT--TTACGATTTAGATTTTATTA	
M67443 Towne	(1860)	CTG---CCCGGAGGACCCCTCCGGAACGGCGCGGACGACG	
AF078102 Rhesus	(1975)	2051	2077
		TGTTTAAATTTTCTTTCTTA---	
M67443 Towne	(1907)	GAGCGGATCGG--GCGGAGCTT	

Figure 42

Figure 42: An alignment of the IFN-gamma nucleotide sequences from human, cat, rodent species.

AF081502 Marmota monax IFN-gamma	(1)	1	50
D30619 Felis catus IFN-gamma	(1)	CTACTGATTTCAACTTCTTT	---
X87308 Homo sapien IFN-gamma	(1)	-----	-----
AF081502 Marmota monax IFN-gamma	(31)	51	100
D30619 Felis catus IFN-gamma	(49)	-----	-----
X87308 Homo sapien IFN-gamma	(1)	-----	-----
AF081502 Marmota monax IFN-gamma	(81)	101	150
D30619 Felis catus IFN-gamma	(99)	-----	-----
X87308 Homo sapien IFN-gamma	(1)	-----	-----
AF081502 Marmota monax IFN-gamma	(131)	151	200
D30619 Felis catus IFN-gamma	(149)	-----	-----
X87308 Homo sapien IFN-gamma	(50)	-----	-----
AF081502 Marmota monax IFN-gamma	(181)	201	250
D30619 Felis catus IFN-gamma	(199)	-----	-----
X87308 Homo sapien IFN-gamma	(100)	-----	-----
AF081502 Marmota monax IFN-gamma	(231)	251	300
D30619 Felis catus IFN-gamma	(249)	-----	-----
X87308 Homo sapien IFN-gamma	(150)	-----	-----
AF081502 Marmota monax IFN-gamma	(279)	301	350
D30619 Felis catus IFN-gamma	(299)	-----	-----
X87308 Homo sapien IFN-gamma	(198)	-----	-----
AF081502 Marmota monax IFN-gamma	(328)	351	400
D30619 Felis catus IFN-gamma	(349)	-----	-----
X87308 Homo sapien IFN-gamma	(247)	-----	-----
AF081502 Marmota monax IFN-gamma	(378)	401	450
D30619 Felis catus IFN-gamma	(399)	-----	-----
X87308 Homo sapien IFN-gamma	(297)	-----	-----
AF081502 Marmota monax IFN-gamma	(428)	451	500
D30619 Felis catus IFN-gamma	(449)	-----	-----
X87308 Homo sapien IFN-gamma	(347)	-----	-----
AF081502 Marmota monax IFN-gamma	(478)	501	550
D30619 Felis catus IFN-gamma	(499)	-----	-----
X87308 Homo sapien IFN-gamma	(397)	-----	-----
AF081502 Marmota monax IFN-gamma	(528)	551	569
D30619 Felis catus IFN-gamma	(549)	-----	-----
X87308 Homo sapien IFN-gamma	(439)	-----	-----

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 August 2000 (10.08.2000)

PCT

(10) International Publication Number
WO 00/46344 A3

- (51) International Patent Classification⁷: C12Q 1/68, C12N 15/09, C07H 21/02, A61K 39/00
- (21) International Application Number: PCT/US00/03086
- (22) International Filing Date: 4 February 2000 (04.02.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/246,178 4 February 1999 (04.02.1999) US
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant: **DIVERSA CORPORATION** [US/US];
10665 Sorrento Valley Road, San Diego, CA 92121 (US).
- Published:
— *With international search report.*
- (72) Inventor: **SHORT, Jay, M.**; 320 Delage Drive, Encinitas, CA 92024 (US).
- (88) Date of publication of the international search report:
28 December 2000
- (74) Agent: **HAILE, Lisa, A.**; Gary Cary Ware & Friedenrich LLP, Suite 1600, 4365 Executive Drive, San Diego, CA 92121-2189 (US).
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 00/46344 A3

(54) Title: NON-STOCHASTIC GENERATION OF GENETIC VACCINES AND ENZYMES

(57) Abstract: This invention provides methods of obtaining novel polynucleotides and encoded polypeptides by use of non-stochastic methods of directed evolution (DirectEvolutionTM). These methods include non-stochastic polynucleotide site-saturation mutagenesis (Gene Site Saturation MutagenesisTM) and non-stochastic polynucleotide reassembly (GeneReassemblyTM). Through use of the claimed methods, genetic vaccines, enzymes, and other desirable molecules can be evolved towards desirable properties. For example, vaccine vectors can be obtained that exhibit increased efficacy for use as genetic vaccines. Vectors obtained by using the methods can have, for example, enhanced antigen expression, increased uptake into a cell, increased stability in a cell, ability to tailor an immune response, and the like. This invention provides methods of obtaining novel enzymes that have optimized physical and/or biological properties. Furthermore, this invention provides methods of obtaining a variety of novel biologically active molecules, in the fields of antibiotics, pharmacotherapeutics, and transgenic traits.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/03086

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C12N 15/09; C07H 21/02; A61K 39/00
US CL : 435/6, 69.3; 536/23.1; 424/184.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 69.3; 536/23.1; 424/184.1; 435/7.6, 69.1, 69.7; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DIALOG, MEDLINE, EMBASE, BIOSIS, US PATENTS, immunomodulatory polynucleotide, library, non-stochastic, mutagenesis, genetic vaccine

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 4,959,312 A (SIROTKIN) 25 September 1990, Abstract.	1-85
Y	CWIRLA et al. Peptides on phage: A vast library of peptides for identifying ligands, Proc. Natl. Acad. Sci. USA. August 1990, Vol. 87, pages 6378-6382, especially page 6378, Materials and Methods.	1-85
Y	HILL et al. Mutagenesis with Degenerate oligonucleotides: An Efficient Method for Saturating a Defined DNA Region with Base Pair Substitutions. Methods in Enzymology. 1987, Vol. 155, pages 558-568, especially pages 561-563.	1-85



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

15 JULY 2000

Date of mailing of the international search report

04 AUG 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

HANKYEL T. PARK, P.H.D.

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/03086

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MURRAY et al. Saturation mutagenesis of a major histocompatibility complex protein domain: Identification of a single conserved amino acid important for allorecognition. Proc. Natl. Acad. Sci. USA. May 1988, Vol. 85, pages 3535-3539, especially pages 3535-3536.	1-85